ENHANCING SOIL SURVEY USING DIFFUSE-REFLECTANCE SPECTROSCOPY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ernest James Neafsey

May 2012

ENHANCING SOIL SURVEY USING PROXIMAL DIFFUSE-REFLECTANCE SPECTROSCOPY

Ernest James Neafsey, Ph. D.

Cornell University 2012

Proximal sensing using visible-near infrared diffuse reflectance spectroscopy (VNIR DRS) has demonstrated substantial potential for the rapid, accurate estimation of key soil properties. Many of these soil properties are diagnostic for the purpose of soil classification and survey.

This study's goals are to: (1) develop a comprehensive VNIR DRS spectral library and PLS1 regression prediction models for a subset of southern New England's benchmark soils and archived soil survey samples; (2) support subaqueous soil mapping efforts in southern New England estuaries and selected salt ponds in Rhode Island by calibrating VNIR DRS as a covariate; (3) determine the utility of using VNIR DRS for measuring sulfidic materials in subaqueous soils, and (4) provide technical assistance to National Cooperative Soil Survey (NCSS) soil survey programs in southern New England.

For benchmark soils, models computed using reflectance spectra measured *in situ* resulted in certain spectral components having root-mean-square errors of prediction (RMSEP) values of 18.89 for % sand, 14.83 for % silt, 5.00 for % clay, 0.75 for pH, 4.95 for cation exchange capacity (CEC), 1.84 for activity and 11.84 for % C; those using first-derivative spectra had RMSEP values of 21.62 for % sand, 20.21 for % silt, 4.80 for % clay, 0.77 for pH, 4.74 for CEC, 1.87 for activity and 10.52 for % C. For subaqueous soils, models computed using reflectance spectra measured from intact, refrigerated soil cores had RMSEP values of 25.81 for % sand, 1355.72 for total sulfide, 22.94 for % SOC, 15.06 for salinity and 0.76 for pH; those using first-derivative spectra had RMSEP values of 26.01 for % sand, 1317.72 for total sulfide,

20.63 for % SOC, 12.17 for salinity and 0.72 for pH. For archived soils, models computed using reflectance spectra measured in the laboratory had RMSEP values of 9909.56 for Fe, 17.57 for % sand, 13.50 for % silt, 6.66 for % clay, 0.70 for pH, 9.34 for CEC, 1.50 for activity and 7.67 for % C; those using first-derivative spectra had RMSEP values of 9418.16 for Fe, 15.61 for % sand, 13.33 for % silt, 5.81 for % clay, 0.77 for pH, 9.46 for CEC, 1.49 for activity and 7.92 for % C. VNIR DRS is recommended for use with benchmark soils, subaqueous soils and archived soils. Realizing the capabilities of VNIR DRS in the soil science community will require substantial cooperation and investment.

Biographical Sketch

Born in Manchester, Connecticut on 20 June 1984, E. J. Neafsey was raised in Stafford Springs, Connecticut and graduated from East Catholic High School in 2001. At Cornell University, he was a Cornell Presidential Research Scholar and conducted research on geostatistical accuracy assessment of digital terrain models (DTMs) created from imagery acquired by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). He received his B.S. degree in January 2005. In addition, he worked for the Connecticut Department of Environmental Protection in Hartford, Connecticut during the summers of 2004 and 2005 in hazard mitigation and permit enforcement. In May 2008, he received his M.S. degree with concentration in Environmental Information Science in the Graduate Field of Soil and Crop Sciences exploring hyperspectral sensing of soil pedons for soil classification and survey.

To my family, with love.

Acknowledgements

This study was supported by USDA-CSREES Federal Formula Funds (HATCH NYC-125401, 125406, 125428); USDA-NRCS Soil Survey Staff- Tolland, CT, Warwick, RI and Taunton, MA; National Soil Survey Center, NRCS-USDA, Lincoln, NE; University of Rhode Island-Bay Campus, Narragansett, RI; the Computational Agriculture Initiative, a joint venture of the Cornell Theory Center and College of Agriculture and Life Sciences and sponsored by USDA-CSREES, Washington, DC; and the Department of Crop and Soil Sciences, Cornell University, Ithaca, NY. In addition, the author would like to thank his special committee, Stephen D. DeGloria, William D. Philpot and Patrick J. Sullivan, as well as his family and friends, for their support in helping him achieve his graduate degree goals.

Table of Contents

Introduction

Rationale

Proximal sensing using visible-near infrared diffuse reflectance spectroscopy's (VNIR DRS) has substantial potential in the rapid, accurate estimation of key soil properties. Many of these soil properties are diagnostic for the purpose of soil classification and survey. For example, numerous soil properties are required to classify a soil. In this case we consider the Charlton series[1] (National Cooperative Soil Survey, 2012).

Soil taxonomic decisions, in terms of classifying a pedon's series and horizon, are made based on documents such as the Official Series Descriptions (OSDs), as well as *Soil Taxonomy* (Soil Survey Staff, 2010). Numerous physical and chemical properties are needed to make these determinations. Setting a "coarse-loamy" class component requires an understanding of soil texture including % sand, % silt, and % clay. The "mixed" class component requires information on mineralogy and pH. The "active" class component relies on a ratio of cation exchange capacity (CEC) to % clay. Each of the horizons described are based on a determination of soil texture (% sand, % silt and % clay), as well as pH. While linguistically descriptive, each technical term has a corresponding numeric definition. Accordingly, classification terminology is derived from quantitative results determined in field and laboratory settings during the soil survey process.

In part to help expedite this effort, the pedometrics (literally "soil measurement") community is developing a body of knowledge linking VNIR DRS to soil characterization. The following (Table 0.1) is an excerpt of spectral analysis predictions just for the soil properties needed to classify a Charlton soil:

[1] The Official Series Description (OSD) is located at https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CHARLTON.html.

Property	Correlation (Range)	Reference(s)
% Sand	0.84	(Cozzolino and Moron, 2003)
% Silt	0.89	(Cozzolino and Moron, 2003)
% Clay	0.93	(Cozzolino and Moron, 2003)
CEC	0.36	(Cozzolino and Moron, 2003)
pH	0.85-0.86	(Reeves and McCarty, 2001; Reeves et al., 1999)

Table I.1 – Correlation values for Charlton series diagnostic soil characterization values

With the exception of CEC, the results for all of these predicted soil properties are well within commonly-accepted standards. In addition, spectroscopy has the advantage of provide rapid results. Well-calibrated VNIR DRS systems can display estimates of desired soil properties within minutes. In contrast, wait-times for wet laboratory results are often six months to two years. While it is up to each user to define his or her data-quality criteria, 84% correlation for % sand, for example, is quite impressive given that the results are immediately available.

Spectroscopy is also cost effective when applied on a large scale. A typical VNIR DRS instrument costs approximately 75 000 USD. The Cornell Nutrient Analysis Laboratory (CNAL) provides high-quality, rapidly delivered soil characterization information to its customers. Recently published prices are 33 USD per sample for their complete soil health assessment chemical test package. Assuming equal labor-costs, the cost of the instrument would be recovered by replacing laboratory analysis of 2273 samples with spectral analysis. If we conservatively assume an average of five samples per pedon, one would have to analyze 455 pedons to achieve price-parity between the two methods. On average, there are 63 counties per state. Seven pedons per county per year is a very reasonable volume demand for soil characterization, considering the various stakeholders (crop and soil scientists, commercial farmers and home gardeners) who request these data.

Goals

The development of this dissertation was guided by the following goals:

1. Establish guidelines for a comprehensive VNIR DRS spectral library and PLS1 regression prediction models for some of southern New England's benchmark soils and archived soil survey samples
2. Support subaqueous soil mapping efforts in southern New England estuaries and selected salt ponds in Rhode Island by calibrating VNIR DRS as a covariate
3. Determine the utility of using VNIR DRS for measuring sulfidic materials in subaqueous soils
4. Provide technical assistance to National Cooperative Soil Survey (NCSS) soil survey programs in southern New England

Overall, the papers contained within this dissertation seek to expand our understanding of how VNIR DRS works within the context of soil survey and of the soils of southern New England. Based on the accuracy, speed and cost data presented above, there are clearly practical benefits of VNIR DRS for pedometrics user groups. However, these benefits will only be realized if they are informed by the science related to its operation and interpretation. It is hoped that this dissertation research makes a substantial contribution to this developing body of knowledge.

References

Cozzolino, D. and A. Moron. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. The Journal of Agricultural Science 140:65-71.

National Cooperative Soil Survey. 2012. Official series description - CHARLTON series. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CHARLTON.html (accessed 28 Jan. 2012).

Reeves, J.B. and G.W. McCarty. 2001. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fibre-optic probe. Journal of Near Infrared Spectroscopy 9:25.

Reeves, J.B., G.W. McCarty and J.J. Meisinger. 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. Journal of Near Infrared Spectroscopy 7:179-194.

Soil Survey Staff. 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75.

Chapter 1: Enhancing stewardship of proximal diffuse reflectance spectral data using the National Cooperative Soil Characterization Database

Abstract

The use of proximal sensing in soil characterization continues to expand. A key area of inquiry is exploring methods of organizing spectral data for archival purposes, efficient retrieval and the development of prediction models for soil characterization properties of interest. The design, operation and maintenance of a national spectral database compatible with National Soil Survey Laboratory (NSSL) standards is proposed. The database proposed has two forms: a distributed database for soil chemometric and other analytical applications and a geodatabase for spatial pedometrics. Significant recommendations include: (1) developing strong international standards for acquiring soil spectra, (2) collecting spectra for all available samples present in the National Cooperative Soil Characterization Database (NCSCD), (3) supporting multiple spectral prediction model types, (4) expanding proximal sensing capabilities, (5) including proximal soil sensing in soil science curricula, and (6) leveraging existing soil survey partnerships. Realizing the promise of proximal diffuse reflectance spectroscopy in the soil science community will require substantial cooperation and investment.

Introduction

An examination of the literature surrounding the use of proximal sensing using visible-near infrared diffuse reflectance spectroscopy (VNIR DRS) as a pedometric tool makes clear the volume of spectral data that has been generated by the soil science community is extensive. Like any scientific research effort, these data have been collected and processed based on specific project objectives. While there

are generally-accepted guidelines for the acquisition of spectral data, based on instrument manufacturers' instructions, literature and specific laboratories' protocols and conventions, the specific procedures used in practice vary widely. For example, some spectra may be collected at 1 nm spectral resolution, while others may be collected at coarser spectral resolution, such as 5-10 nm. These choices are based upon the practitioner's best understanding of the spectral response of the soil properties being examined. Additionally, there are many configurations (mug lamp, contact probe, bare fiber, pistol grip, etc.) of the sensor probe, the specific geometries of which will affect the spectra they collect. More critically, physical properties of the soil samples (clod size, moisture content, etc.) vary substantially. There remains a strong tradition in the proximal diffuse reflectance spectroscopy community of collecting spectra from air-dried, ground, and sieved soil samples. However, there is an increasing number of studies which examine archived and/or field moist soil samples, some of the latter are measured *in situ* in the soil profile. These and other factors can add substantial variance terms in the set of collected spectra.

Several papers have proposed and/or implemented database designs which meet the need to encompass a wide range of methodologies. In addition, efforts have been made at universities to compile their spectral holdings (Dunn et al., 2002, Shepherd and Walsh, 2002; Brown et al., 2006; Viscarra Rossel et al., 2008). These institutions may be seeking maximum leverage from existing soil spectral and characterization data to avoid the costs associated with unnecessary duplicative sampling. Additionally, proximal diffuse reflectance spectroscopy may be used by different units and research groups within an organization. Here communication is enhanced and advanced collaboration is possible. Given the high initial investment in spectroscopic instrumentation, an institutional census of spectral samples can illustrate the relative low cost of each spectrum. However, no centralized repository

(international or national) has been developed to store and develop the potential of the soil research community's spectral resources.

Instrument calibration for key soil properties will be critical for developing proximal diffuse reflectance spectroscopy as a field tool. This will require the sharing of spectra and corresponding soil characterization data for the computation of spectral prediction models (Brown et al., 2006). Soil spectra and characterization data are collected on a weekly basis. It will be important for our spectral database and prediction model infrastructure to be up to date with current research, so that all users may benefit. The objective of this study is to propose the design, operation and maintenance of a national spectral database, compatible with National Soil Survey Laboratory (NSSL) standards.

Approach

Despite the broad interest in soil spectral databases and libraries, there has been only periodic treatment of those which incorporate soil properties over a geographically diverse area. Potential reasons include the substantial temporal and financial investments required to carry out database construction at the proper scale and magnitude. Nevertheless, the soil science community is trending towards the adoption of standards for soil spectral databases and their eventual publication. This section explores a selection of those that have been proposed and developed to date (Table 1.1).

Dunn et al. (2002) describe a spectral database constructed for the Riverine Plain soils of New South Wales in Australia. The samples in the database are classified by depth class, rather than horizon. It uses "550 topsoil (0-10 cm) and 300 subsoil (40-50 cm) samples" (Dunn 2002). Typical of proximal diffuse reflectance spectroscopy samples they "were air dried and ground before scanning" (Dunn 2002). Partial Least Squares (PLS) regression results "demonstrated that NIRS can successfully determine some soil properties in both the topsoil and subsoil. In the topsoil, cation exchange capacity (CEC), exchangeable Ca and Mg, pH, and Ca : Mg ratio were predicted with a high level of accuracy and organic carbon and

exchangeable sodium percentage (ESP) with an acceptable level of accuracy. In the subsoil, CEC, exchangeable Na, Ca, Mg, ESP, pH and Ca : Mg ratio was all predicted with a high degree of accuracy" (Dunn 2002).

Shepherd and Walsh (2002) describe a spectral database constructed for eastern and southern Africa. Like Dunn et al. (2002), the samples in the database are classified by depth class, rather than horizon. It uses 1000 "topsoil (0- to 15- or 0- to 20-- cm depth) samples," which "were air-dried, passed through a 2-mm sieve, and stored in paper bags at room temperature" (Shepherd and Walsh 2002). Multivariate adaptive regression splines (MARS) yielded the following R^2 values for selected soil properties: "exchangeable Ca, 0.88; effective cation exchange capacity (ECEC), 0.88; exchangeable Mg, 0.81; organic C concentration, 0.80; clay content, 0.80; sand content, 0.76; and soil pH, 0.70" (2002). Spectral data contained in this database have a spectral resolution of 10 nm.

Brown et al. (2006) describe a spectral database constructed using "3768 samples from all 50 U.S. states and two tropical territories and an additional 416 samples from 36 different countries in Africa (125), Asia (104), the Americas (75) and Europe (112). The samples were selected from the National Soil Survey Center archives in Lincoln, NE, USA, with only one sample per pedon and a weighted random sampling to maximize compositional diversity." The samples were horizon based, with the following distribution: "5 O, 1106 A, 70 E, 2412 B, 518 C and 4 R horizons with 69 miscellaneous materials" (Brown et al. 2006). Samples were air dried and had a particle size of less than 2 mm. Using "auxiliary predictors including sand content or pH, we obtained validation root mean squared deviation (RMSD) estimates of 54 g kg^{-1} for clay, 7.9 g $kg^{-1}$for soil organic C (SOC), 5.6 g kg^{-1} for inorganic C (IC), 8.9 g kg^{-1} for dithionate–citrate extractable Fe (FEd), and 5.5 $cmol_c$ kg^{-1} for cation exchange capacity (CEC) with NH_4 at pH = 7" (Brown et al. 2006). Boosted regression trees (BRT) were used for prediction model calculation. Further, they "anticipate that calibrations sufficient for many applications might be obtained with large but

obtainable soil-spectral libraries (perhaps 10^4–10^5 samples)" (Brown et al. 2006). Spectral data contained in this database have a spectral resolution of 10 nm.

Viscarra Rossel et al. (2008) describe a spectral database constructed using 213 samples "from the Upper Namoi, Namoi, and Gwydir Valley catchments of north-western New South Wales (NSW) and the McIntyre region of southern Queensland (Qld)" (Viscarra Rossel et al. 2008). "The soil was sampled from different layers, including the 0–0.10, 0.10–0.20, 0.30–0.40, 0.60–0.70, and 0.70–0.80 m. They were stored in sealed plastic containers as ground samples with a size fraction of ≤2 mm" (2008). PLS predictions models "for organic C (OC), cation exchange capacity (CEC), clay content, exchangeable Ca, total N (TN), total C (TC), gravimetric moisture content θ_g, total sand and exchangeable Mg were robust and produced accurate results ($R^2_{adj.}$ > 0.75 for both cross and test set validations)" (Viscarra Rossel et al. 2008). Spectral data contained in this database have a spectral resolution of approximately 2 nm (8 cm^{-1}).

Philosophically, Viscarra Rossel et al. (2008) indicate the desirable elements of a spectral database:

"Three important requirements for the development of a soil spectral library are: (i) it should contain as many samples as are needed to adequately describe the soil variability in the region in which the library is to be used; (ii) the samples should be carefully subsampled, handled, prepared, stored, and scanned (everything that has happened to the sample up to the time of scanning will be embodied in the sample and recorded in the spectra); and (iii) the reference soil analytical data used in the calibrations should be acquired using reliable and accredited analytical procedures ."

The International Soil Reference and Information Centre (ISRIC) and the World Agroforestry Center (ICRAF) describe a spectral database constructed from "785 soil profiles (4,437 samples) selected from the ISRIC Soil Information System (ISIS)" (ICRAF-ISRIC 2012). "The samples are from 58 countries

spanning Africa, Asia, Europe, North America, and South America" (ICRAF-ISRIC 2012). Samples were air dried and sieved to a particle size of less than 2 mm. As this database is primarily designed as a resource for the soil science community, soil characterization data are stored with it, but no prediction models are provided or suggested. Spectral data contained in this database have a spectral resolution of 10 nm.

In summary (Table 1.1), spectral databases have varied in size and sampling unit based upon location. State or continental region based databases tend to have higher sampling densities and be based upon layers instead of horizons. These schemes support the soil modeling needs of these regions, while global databases must be interoperable with horizon data, which remains the most common soil sampling unit. All samples are air-dried and use the fraction 2 mm or smaller. While this fraction is used for most soil analytical techniques, expanding these databases to include field samples would be beneficial. PLS regression is used in three of the four databases surveyed which include prediction models, with one using BRT in conjunction with PLS and the final database using MARS. Four of the five databases have a spectral resolution of 10 nm, while the final database has a spectral resolution of 2 nm.

Citation	Location	*N*	Layer/Horizon?	Sample Type	Prediction Model Used	Spectral Resolution
(Dunn et al., 2002)	NSW, AUS	550 topsoil 300 subsoil	Layer	Air-dried, ≤ 2mm	PLS	10 nm
(Shepherd and Walsh, 2002)	E and S Africa	1000	Layer	Air-dried, ≤ 2mm	MARS	10 nm
(Brown et al., 2006)	Global	3768	Horizon	Air-dried, ≤ 2mm	PLS/BRT	10 nm
(Viscarra Rossel et al., 2008)	NW NSW and S QLD, AUS	213	Layer	Air-dried, ≤ 2mm	PLS	2 nm
(ICRAF-ISRIC, 2012)	Global	4437	Horizon	Air-dried, ≤ 2mm	N/A	10 nm

Table 1.1 – Summary of spectral databases

Proposed Database Design

The proposed spectral database must serve two general audiences: 1) those that require site specific soil characterization data for a particular application (soil survey, farm management, field engineering, etc.) and 2) the soil modeling community. As these audiences are substantially interlinked it is important that database components be linked. Additionally, the database design needs to be able to accommodate the diversity of spectra collected. Proposed here are a database (Figures 1.1 and 1.2) for computational and predictive purposes and a geodatabase (Figure 1.3) for applications where geospatial areas of interest are important.

The database design illustrated below (Figures 1.1 – 1.2) proposes to efficiently deliver spectra to users and leverage the power of the NSSL's NCSCD (National Cooperative Soil Survey, 2012). Of immediate note is support for spectra collected using different, accepted protocols. These may include

combinations of spectra collected on field-moist or air-dried samples, in the field or laboratory and probe type (mug lamp, contact probe, bare fiber, foreoptics, etc.).

All spectral samples are linked to the NCSCD (which contains the full description of the soil sample) via the natural_key field in the Layer table (Figure 1.1). Figure 1.2 provides the general tables in which the spectral samples will reside. The samples are labeled by natural_key and the fields represent wavelengths, in increments of 10 nm, from 350 nm to 2500 nm. A spectral resolution of 10 nm is consistent with the literature and is within the 255 field limit for tables in Microsoft Access 2010. Since the NSSL is "stor[ing] the raw data to create maximum output flexibility," (Reinsch, 2010) the database's spectral resolution can be increased at a later time (presuming Access' field limit is increased in the future or the NSSL transitions to another database platform).

Another major feature of the design is that users can select reflectance values for specified wavelengths, rather than having to download complete spectra. Based on the literature, many practitioners use only a portion of the spectra for prediction model computation purposes. Furthermore, this approach saves valuable server space by eliminating the need for wavelengths to be stored explicitly in the database.

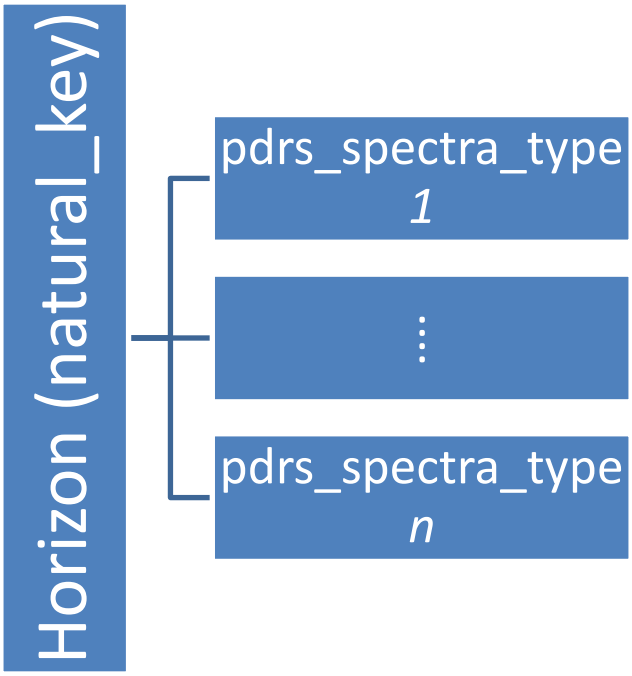


Figure 1.1 - Proposed NSSL spectral database (general form)

natural_key	W350	W360	…	W2490	W2500
00P00001					
⋮					
09N01866					

Figure 1.2 – Proposed NSSL spectral database (table for each pdrs_spectra_type)

Figure 1.3 illustrates a proposed geodatabase where the primary hierarchy is a pedon list, linked to the NSSL soil characterization database by way of the natural key used for all pedons. For ease of recognition, the geodatabase should be labeled by user pedon ID. Important pedon and sample attributes (including percent carbon, geographic location, year, etc.) can be accessed through the main NSSL soil characterization database for spatial analysis purposes to reduce storage requirements. The geodatabase design implies each pedon has a known, stored location. Spectral types, used to classify acquisition modes and conditions, are in a tertiary hierarchy to encourage building spectral sets for analysis which were collected under similar conditions. Queries can be developed to aggregate spectra by desired type. The individual spectra by horizon designation, stored in the appropriate format, follow in a quaternary hierarchy.

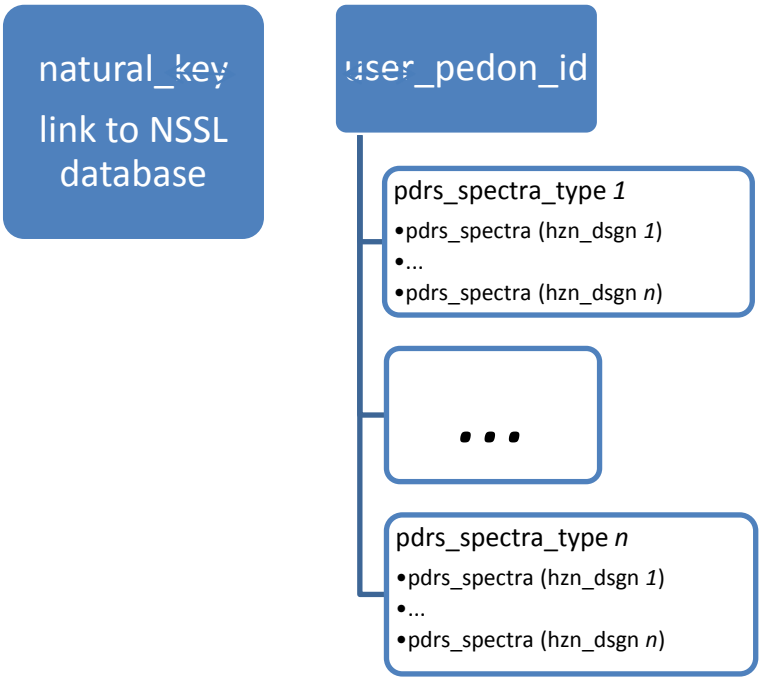


Figure 1.3 – Proposed geodatabase (pedon list)

Scripts can be used to pull spectra from the main NSSL spectral database as it is updated. To accommodate the *GlobalSoilMap.net* project, database implementations can also be designed to include spectra by soil layer (typically 50 mm increments) in addition to by horizon.

Perhaps the most powerful aspect of the proposed database is the ability to combine spectral data with the richness of the NCSCD to develop prediction models for a given application. These custom spectral libraries can be downloaded in popular formats (ASCII text, Microsoft Office Excel, The Unscrambler, etc.). The proposed database update cycle in Figure 1.4 envisions additions to both soil characterization and spectral data through continuing sampling programs improving the quality of updated prediction models. The model estimates update cycle can largely be automated through the use of open-source statistical scripting languages (such as R) to produce updated prediction models and deliver them to users in their desired format. The update process should also have some provision for maintaining

statistical quality, such as identification of outliers and grouping samples appropriately. Much of this can be automated with some measure of human supervision.

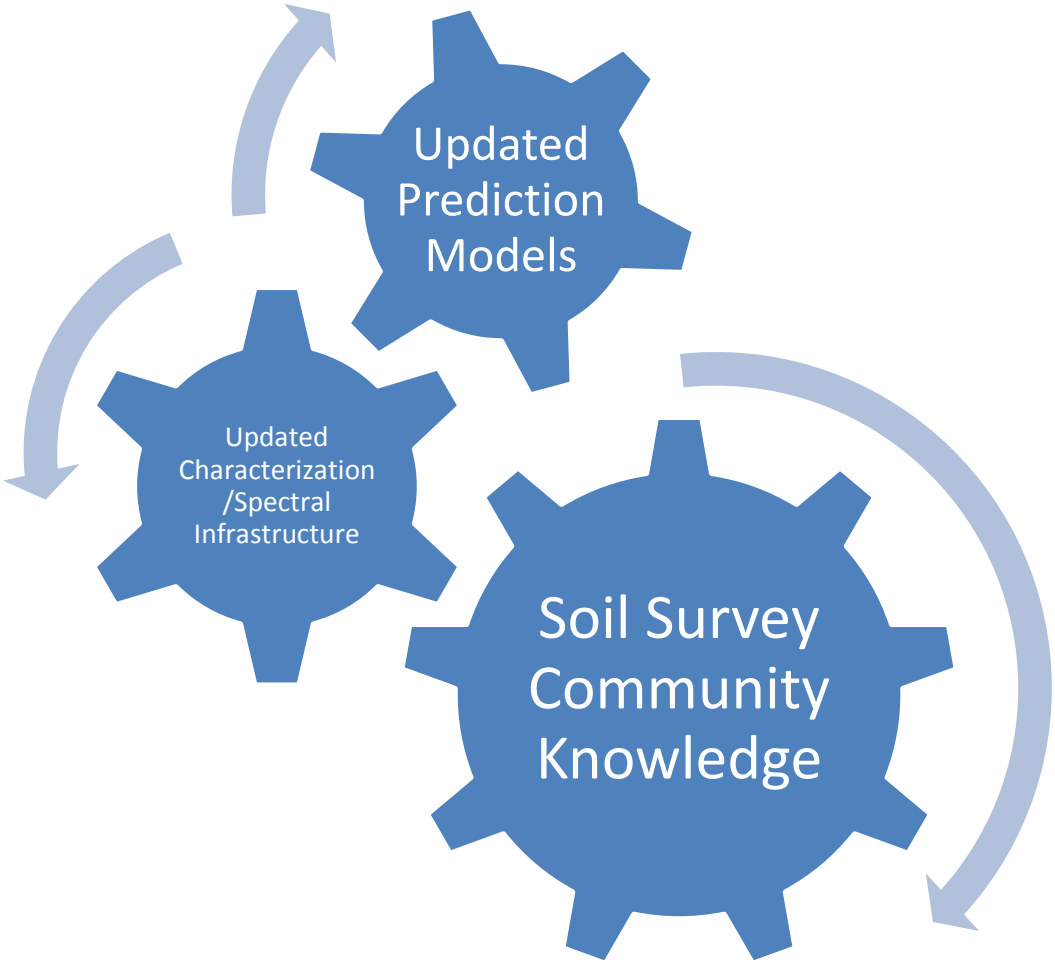


Figure 1.4 - Proposed update cycle

In the future, these prediction models might be loaded onto field spectrometers (perhaps using The Unscrambler Prediction Engine[2]) to help produce on-site estimates of desired soil properties. Updated prediction models could then be transmitted to field spectrometers via mobile broadband or another appropriate Internet connection. Proposed applications include aiding soil survey interpretations,

[2] <http://www.camo.com/rt/Products/Unscrambler/olup.html>

assessment of agricultural inputs, engineering, and site assessment. As the updated prediction models require traditional laboratory methods to compute revised prediction models, the system envisioned here is not intended to fully replace conventional methods.

Recommendations

Fully leveraging VNIR DRS in the soil science community will require a framework for the collection, organization, maintenance and retrieval of spectral holdings and associated soil property prediction models. The following recommendations are made to suggest a path forward for the adoption of the technology in the field.

Recommendation 1: Develop strong international standards for acquiring soil spectra.

Practitioners must always have flexibility in the design and execution of data collection campaigns. The vibrancy and diversity of this sector of the pedometrics community is a testament to its relevance and growth potential. However, much as conventional soil analytical methods have been standardized, there is a need to harmonize soil spectral collection procedures. This discussion, including all potential users, should include the following:

- Instrument calibration: How is the radiometric accuracy of the instrument best maintained throughout the day? Its lifetime? How do we ensure spectral and radiometric (different light sources, fields of view, incidence/viewing angles, etc.) compatibility among data collected using different instruments and vendors?
- Soil sample standards: As a matter of scientific definition, what constitutes “field moist?” How can this be interpreted practically in the field? What are the procedures for collecting spectra for samples stored in contact with the atmosphere? What is an acceptable variance in moisture for archived samples?

- Field conditions: Spectra may be collected in soil survey pits, subaqueous vibracores or soil samples extracted using augers. What is the protocol for collection under these conditions?
- Collection parameters: What is the most flexible form of raw spectra? How many replicates are needed for each sample? What is the desired spectral resolution and waveband?
- Publication standards: What is the best resampling (and possibly smoothing methods, including derivatives or splines) method for spectral publication? What is the desired waveband?
- Combination with other proximal/remote sensing methods: Can proximal diffuse reflectance spectroscopy be used with mid-infrared (MIR) spectra? What is the preferred method for calibrating remotely sensed hyperspectral imagery using proximally sensed spectra?

The answers to these questions do not present themselves easily, but must be pursued in cooperation with stakeholders.

Recommendation 2: Collect spectra for all available samples present in the NCSCD.

The proposed spectral database design will require a decadal investment in both conventional soil sampling and spectral data collection. There are 227 215 records (and corresponding samples) in the NCSCD as of 2010. This is just over twice the top of the range of 100 000 samples recommended by Brown et al. (2006) for “many applications.” This suggests that, even using half of available data, given the possibility missing data in the NCSCD for some properties, a near-universal spectral database is within reach. To accomplish this, the NSSL should collect spectra for all available samples present in the NCSCD.

Recommendation 3: Support multiple spectral prediction model types.

While there are certainly popular options for the development of spectral prediction model types, there is yet no strong consensus among the pedometrics community as to which should enjoy universal

adoption. In fact, this debate remains an important one to the field of proximal soil sensing. Therefore, until recent literature sees equilibrium in the prediction model type used, the BRT, MARS and PLS model types should be made available to users. Some guidance should be given as to the assumptions, operation, advantages and limitations of each prediction model type.

Recommendation 4: Expand proximal sensing capacity.

Fulfilling Recommendation 2 in a timely manner will require the provision of training for soil survey staff, proximal sensing specialists and instrumentation to collect spectra at the NSSL, Natural Resources Conservation Service (NRCS) state offices, or Major Land Resource Area (MLRA) regional offices. Two to three day workshops just before or after biennial regional National Cooperative Soil Survey (NCSS) conferences could cover proximal soil sensing theory, instrument operation and spectral processing and interpretation. People are the most expensive part of this recommendation, but their experience gained through the process will be invaluable for the continued development and interpretation of spectral information.

Recommendation 5: Include proximal soil sensing in soil science curricula.

Just as soil science programs provide training in traditional analytical procedures, the curricula of participating educational institutions should begin to offer courses in proximal soil sensing. These courses should include information on sampling design, relevant soil physics and proximal sensing theory, instrument operation, data management, statistical analysis and interpretation. As alluded to in Recommendation 4, a trained workforce is the best way to ensure the efficacy of proximal soil sensing. Additionally, this will reduce future employer training costs and needs.

Recommendation 6: Leverage existing soil survey partnerships.

In conclusion, realizing the promise of proximal diffuse reflectance spectroscopy in the soil science community will require substantial cooperation and investment. Science is a social enterprise, and progress is made only through the interactions of its participants. The meetings of the International Union of Soil Scientists, the Soil Science Society of America, the American Society for Photogrammetry and Remote Sensing, Pedometrics and the National Cooperative Soil Survey provide ample opportunities for the exchange of ideas regarding proximal soil sensing. The flexibility and broad interest base of each of the above organizations opens the possibility of for effective communication and interactions on the subject.

References

Brown, D.J., K.D. Shepherd, M.G. Walsh, M. Dewayne Mays and T.G. Reinsch. 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132:273-290.

Dunn, B.W., H.G. Beecher, G.D. Batten and S. Ciavarella. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis-a case study from the riverine plain of south-eastern Australia. Aust. J. Exp. Agric. 42:607-614.

ICRAF-ISRIC. 2012. A globally distributed soil spectral library: Visible near infrared diffuse reflectance spectra. http://www.africasoils.net/data/ICRAF-ISRICspectra (accessed 28 Jan. 2012).

National Cooperative Soil Survey. 2012. National cooperative soil characterization database. http://ssldata.nrcs.usda.gov/ (accessed 28 Jan. 2012).

Reinsch, T.G. 2010. Soil NIR DRS spectral database.

Shepherd, K.D. and M.G. Walsh. 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J. 66:988.

Viscarra Rossel, R.A., Y.S. Jeon, I.O.A. Odeh and A.B. McBratney. 2008. Using a legacy soil sample to develop a mid-IR spectral library. Aust. J. Soil Res. 46:1-16.

Chapter 2: An assessment of the feasibility of using in situ proximal diffuse reflectance spectroscopy in the characterization of benchmark soils in southern New England

Abstract

Previous literature suggests that proximal diffuse reflectance spectroscopy is effective in the estimation of soil characterization properties in a laboratory setting. However, little research has been done to assess this method's feasibility *in situ* during conventional benchmark soil survey operations. Specific study objectives are to: (1) assess the feasibility of using in situ proximal diffuse reflectance spectroscopy in the characterization of benchmark soils in southern New England and (2) make recommendations as to proximal diffuse reflectance spectroscopy's use in NRCS field benchmark soil survey operations. This research seeks to test the technology's previously demonstrated efficacy in estimating soil properties (using PLS1 regression), useful for soil classification and survey, in a field setting in southern New England. PLS1 regression models computed using reflectance spectra had RMSEP values of 18.89 for % sand, 14.83 for % silt, 5.00 for % clay, 0.75 for pH, 4.95 for cation exchange capacity (CEC), 1.84 for activity and 11.84 for % C; those using first-derivative spectra had RMSEP values of 21.62 for % sand, 20.21 for % silt, 4.80 for % clay, 0.77 for pH, 4.74 for CEC, 1.87 for activity and 10.52 for % C. Based on these RMSEP values, proximal diffuse reflectance spectroscopy is recommended for use with conventional benchmark soil survey operations.

Introduction

The inventory, monitoring and management of benchmark soils will require timely and accurate information. The features of benchmark soils and the requirements for proximal diffuse reflectance spectroscopy to be effective are complementary. Viscarra Rossel et al. (2008) state that spectral libraries "should contain as many samples as are needed to adequately describe the soil variability in the region in which the library is to be used" (Viscarra Rossel et al. 2008). A data policy which prioritizes collecting spectra for benchmark soils as they are surveyed has the potential to deliver maximum return on investment (Giasson et al., 2000).

The Natural Resources Conservation Service's *Soil Survey Handbook* defines benchmark soils as being "one of large extent within one or more major land resource areas (MLRA), one that holds a key position in the soil classification system, one for which there is a large amount of data, one that has special importance to one or more significant land uses, or one that is of significant ecological importance" (U.S. Department of Agriculture, Natural Resources Conservation Service, 2012). "Benchmark soils, while being important soils in their own right, are also intended to serve as proxies for other similar soils. Their purpose is to focus data collection and the investigative effort on soils that have the greatest potential for extending collected data and resultant interpretations to other soils" (U.S. Department of Agriculture, Natural Resources Conservation Service, 2012).

The goal of this study is to test the technology's previously demonstrated efficacy in estimating soil properties, useful for soil classification and survey, in the field setting in southern New England. These include texture (% sand, % silt and % clay), pH, cation exchange capacity (CEC), clay activity and percent Carbon (% C). Previous literature suggests that proximal diffuse reflectance spectroscopy is effective in the estimation of soil characterization properties. Viscarra Rossel et al. (2006) report R^2 values of 0.59 for % sand, 0.41 for % silt, 0.60 for % clay, 0.57 for pH and 0.13 for CEC (Viscarra Rossel et al., 2006).

Walvoort and McBratney reported a low root mean square error (RMSE) of 0.06 for % C (Walvoort and McBratney 2001). Estimates of activity (ratio of cation exchange capacity to percent clay content) have yet to be recorded in the literature, but its proportionality to clay content implies similar values. However, little research has been done to assess this method's feasibility *in situ* during conventional soil survey operations (Neafsey, 2008). Specific objectives of this study are to: (1) assess the feasibility of using in situ proximal diffuse reflectance spectroscopy in the characterization of benchmark soils in southern New England and (2) make recommendations as to proximal diffuse reflectance spectroscopy's use in NRCS field benchmark soil survey operations.

Materials and methods

Study area

The project study area is located within the Northeastern Forage and Forest Land Resource Region. Topography is characterized by plateaus, plains and mountains. The climate is cool and humid, experiencing an average precipitation of 865 – 1575 mm per year. Average annual temperature is 4 – 9 °C. There are 130 – 200 freeze days, with 110 in the highest elevations and 240 in some coastal areas. Total withdrawals of freshwater from the region are, on average, 72 ML/d, 93 % of which is from surface water and 7 % from groundwater. Timber and manufacturing interests consume approximately 78 % of this total, with the rest for public supply.

Soils in this region are primarily Entisols or Spodosols, commonly with a fragipan. In areas with limey geology, Alfisols occur with a fragipan. Higher elevations are characterized by Ochrept and Orthods, while lower regions feature Aqualfs, Aquepts and Histosols. Fluvents in floodplains are limited in extent but are locally important. The temperature regimes are frigid or mesic, udic in terms of soil moisture and exhibit a mixed mineralogy.

The primary land cover is forest (especially in areas of rugged relief), with 98 % of the land in private ownership. Logging, for timber and pulpwood, is a commercially important activity, as is the production of Christmas trees and maple syrup. The region grows forage and grains to feed local dairy stock, however local conditions can allow for the production of food crops for human consumption. Wildlife habitat and recreation are major land uses, especially where stoniness and steep slope has precluded other uses (USDA, 2006).

DRS measurements

Soil pits were excavated to provide access for soil profile description, sampling, and spectral data acquisition. While the soil pits varied slightly, typical dimensions were 2500 mm (length) by 500 mm (width) by 1750 mm (depth). These dimensions are larger than most soil pits and were designed to accommodate the ASD FieldSpec Pro FR spectrometer (Analytical Spectral Devices, Inc., Boulder, CO[3]) and two or three people needed for soil spectra collection. The spectrometer collects data at a spectral resolution of 1 nm. It is also a three-spectrometer system, with detectors for the VNIR (350 – 1000 nm), SWIR1 (1001 – 1780 nm) and SWIR2 (1781 – 2500 nm) spectral regions. Discontinuities can be observed (Figure 2.3) in the spectra in transitioning between these regions. The pedon face was then cleaned to reduce any soil mixing effects caused by the excavation equipment. While the soil profile was being prepared for description and sampling, the spectrometer was prepared to acquire spectral data. A dark current (zero reflectance) was measured for calibration purposes. The instrument was also calibrated to a white reference (spectralon) before each soil profile was scanned.

Soil surveyors typically analyze soil pedons by horizon, so most spectral data in this study were collected by horizon as well. To accomplish this, NRCS soil surveyors physically marked all soil horizon boundaries. To mitigate any periodicity in the soil profile, a random starting point between 0 - 50 mm from the left

[3] <http://www.asdi.com>

side of the excavated and prepared pedon was chosen near the middle of each horizon in the vertical dimension.

Three sets of five spectra (each an instantaneous average of ten spectra) were collected for each horizon, with the random starting point and the second and third points located 100 and 200 mm away from the first point, respectively. Spectra were collected in full contact with the soil profile (see Figures 2.2 and 2.3). A clean, dry washcloth was used to wipe away any residue left on the contact probe lens after each sample was collected. In addition, a white reference (full reflectance, contact probe model) was taken after each soil horizon to keep the instrument calibrated throughout the scanning of each soil pedon.

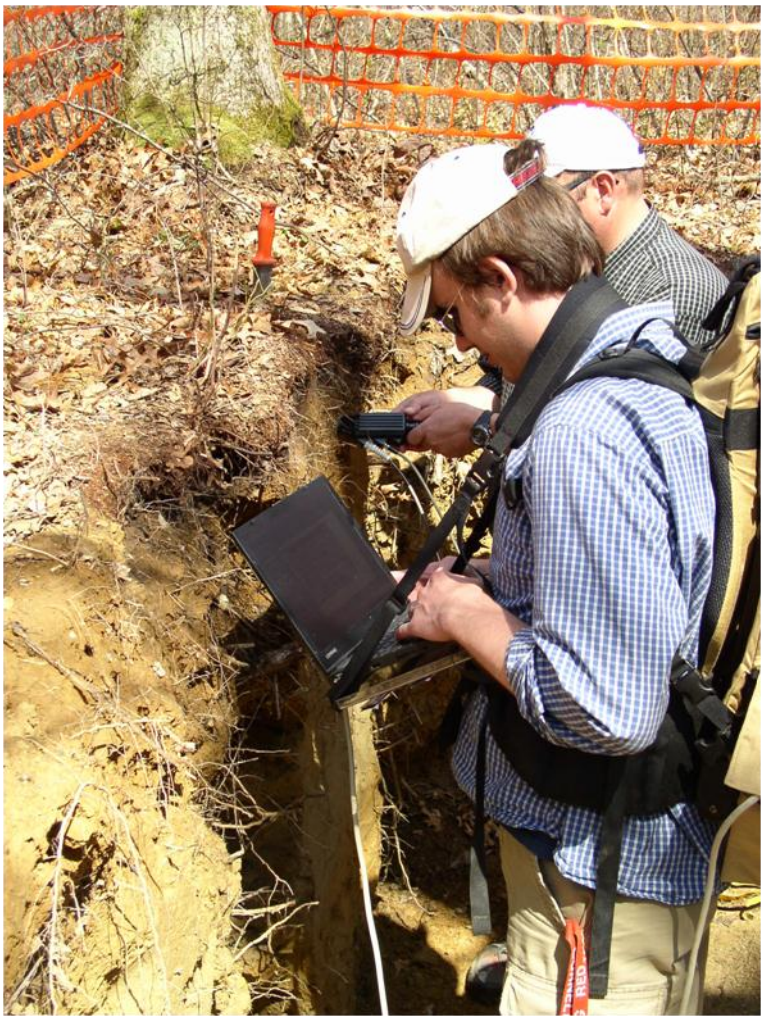


Figure 2.1 – Field instrument setup – side

Figure 2.2 – Field instrument setup – pedon face

Sampling for laboratory analysis

After profile description and spectral acquisition were complete, the Natural Resources Conservation Service (NRCS) soil survey team collected soil samples for the pedon for submission to the National Soil Survey Laboratory (NSSL) at the National Soil Survey Center (NSSC) in Lincoln, NE. All soil samples were analyzed by the NSSL to determine their physical, chemical, and mineralogical properties according to NRCS soil survey guidelines (U.S. Department of Agriculture, Natural Resources Conservation Service, 2012; National Cooperative Soil Survey, 2012).

Data analysis

To facilitate data analysis, the instrument native files (remote sensing reflectance calculated by the ASD software) for each horizon were imported into The Unscrambler software program,[4] a multivariate statistics package which specializes in spectral processing and analysis. Reflectance spectra can also be imported to universally-readable ASCII files by ASD's ViewSpecPro, but direct importation provides a more efficient workflow and ensures the preservation of variable names.

The spectra used for analysis for each horizon represent an average of 15 original spectra collected at the three points for each horizon, with five spectra collected at each point. Averaging accomplishes instrument noise reduction and accounts for variance in the samples.

First derivatives were calculated using the difference between the reflectance values associated with adjacent wavelengths at a resolution of 1 nm to preserve full spectral information. As soil moisture can reduce the amplitude of soil spectra, derivative analysis is used to normalize these amplitudes, highlighting the constituents of interest (ASD Inc., 2012).

PLS1 regression modeling

Partial least squares regression was used to evaluate the statistical contributions of many X-variables (predictors) to one Y-variable (responses). In this study, the reflectance values for the waveband 350 – 2500 nm comprised the set of X-variables. Excluded due to falloff in detector sensitivity were the wavebands 350 – 450 nm and 2400 – 2500 nm (marked in gray in Figure 2.3). Excluded also were water absorption wavebands 1350 – 1500 nm and 1850 – 2000 nm (marked in blue in Figure 2.3). Selected physical and chemical soil properties were the Y-variables. A common statistical tool in soil chemometrics, PLS1 regression allows the user to determine which spectral wavelengths are correlated with properties of interest (Vasques et al., 2008; Vasques and Sickman, 2009; Viscarra Rossel, 2008).

[4] <http://www.camo.com/rt/Products/Unscrambler/unscrambler.html>

Results and discussion

Sample descriptions

In 2008, NRCS benchmark soil survey efforts focused on southeastern New England sulfaquents, dystrudepts and sulfihemists (Table 2.1). These soils were located in estuarine and riverine environments, as well as other sandy locations. In particular, the Ipswich and Westbrook soils are subject to inundation and were sampled using a McAuley auger (as opposed to the soil pits used to sample the other soils) for soils with high content of peat. McAuley augers confine the soil column segment being extracted to a compartment, maintaining its structure.

According to the median (Table 2.2), the soils in this study are characterized by high silt content, with relatively low sand and clay contents. They are also very strongly acid, have a CEC outside of general agricultural limits and are slightly active. Carbon content is also relatively low.

There are several features of particular interest in the sample spectral set (Figure 2.3) for a single pedon representing the Newport series (natural_key 08N0528), reflectance and first-derivative. The primary variable controlling spectral reflectance is sand content. The horizon with the highest maximum reflectance is the Bw, with 64 % sand, while the lowest reflectance appears in the 2Cd3 horizon, with 45 % sand. Major water absorption bands (marked in blue) are located at 1350 – 1500 nm (relatively shallow, despite the moist field conditions reducing maximum reflectance) as well as 1850 – 2000 nm. The changes in spectral shape in the visible waveband (400 – 700 nm) demonstrate the differences in soil color. It is particularly noticeable that the maximum reflectance occurs in the visible waveband.

A derivative can be effective in locating an extremum since the derivative will be transitioning through zero at that point (a “zero-crossing”). The derivative is sensitive to the slope rather than the magnitude, and can be very sensitive to subtle features. However, the subtle features can easily be obscured by noise. Here, the zero-crossing corresponding to the maxima of the reflectance spectra all occur in the

region outlined in a red ellipse. The noise in the derivative obscures the specific position of the crossing. A reliance on PLS1 regression avoids the need to arbitrarily alter the spectra by smoothing and allows the technique to tease out significant responses. Note that the noise increases at the wings of the detectors, i.e., the noise is greater in the UV and SWIR ends of the spectrum, and also increases closer to the transition between spectrometers. The abrupt increase in noise is particularly apparent in going from the SWIR1 to the SWIR2 spectrometer.

Pedon ID	Taxon Name	Class Name
S08MA023001	Ipswich	Fine-silty, mixed, superactive, acid, mesic Typic Sulfaquent
S08MA023002	Merrimac	Sandy, mixed, mesic Typic Dystrudept
S08MA023003	Merrimac	Sandy, mixed, mesic Typic Dystrudept
S08MA023004	Winooski	Coarse-silty, mixed, superactive, mesic Fluvaquentic Dystrudept
S08CT011001	Ipswich	Euic, mesic Typic Sulfihemist
S08CT011002	Westbrook	Loamy, mixed, euic, mesic Terric Sulfihemist
S08RI005-001	Newport	Coarse-loamy, mixed, active, mesic Typic Dystrudept
S08RI009-020	Bridgehampton	Coarse-silty, mixed, active, mesic Typic Dystrudept

Table 2.1 – Sample descriptions

	Field - % Sand	Field - % Silt	Field - % Clay	Field - pH	Field - CEC	Field - Activity	Field - % C
Units	%	%	%	pH	meq+/100g	scalar	%
#	42	42	39	46	27	29	51
Mean	41.49	50.01	9.15	4.95	6.68	1.46	9.06
Max	98.20	87.20	27.60	6.40	28.49	9.89	48.57
Min	3.30	1.80	0.10	3.40	0.50	0.28	0.02
Range	94.90	85.40	27.50	3.00	27.99	9.61	48.55
Std Deviation	31.09	26.83	7.41	0.82	7.04	1.94	13.49
Variance	966.60	719.89	54.96	0.67	49.57	3.77	182.01
RMS	51.63	56.60	11.72	5.02	9.61	2.40	16.14
Skewness	0.47	-0.40	1.15	-0.17	1.99	3.41	1.62
Kurtosis	-1.33	-1.30	0.05	-0.48	3.37	13.04	1.48
Median	26.55	58.85	7.00	5.00	4.90	0.76	1.23
Q1	14.03	24.13	4.05	4.50	2.67	0.56	0.28
Q3	71.73	74.05	9.25	5.45	6.50	1.29	10.99

Table 2.2 – Soil characterization properties summary statistics table

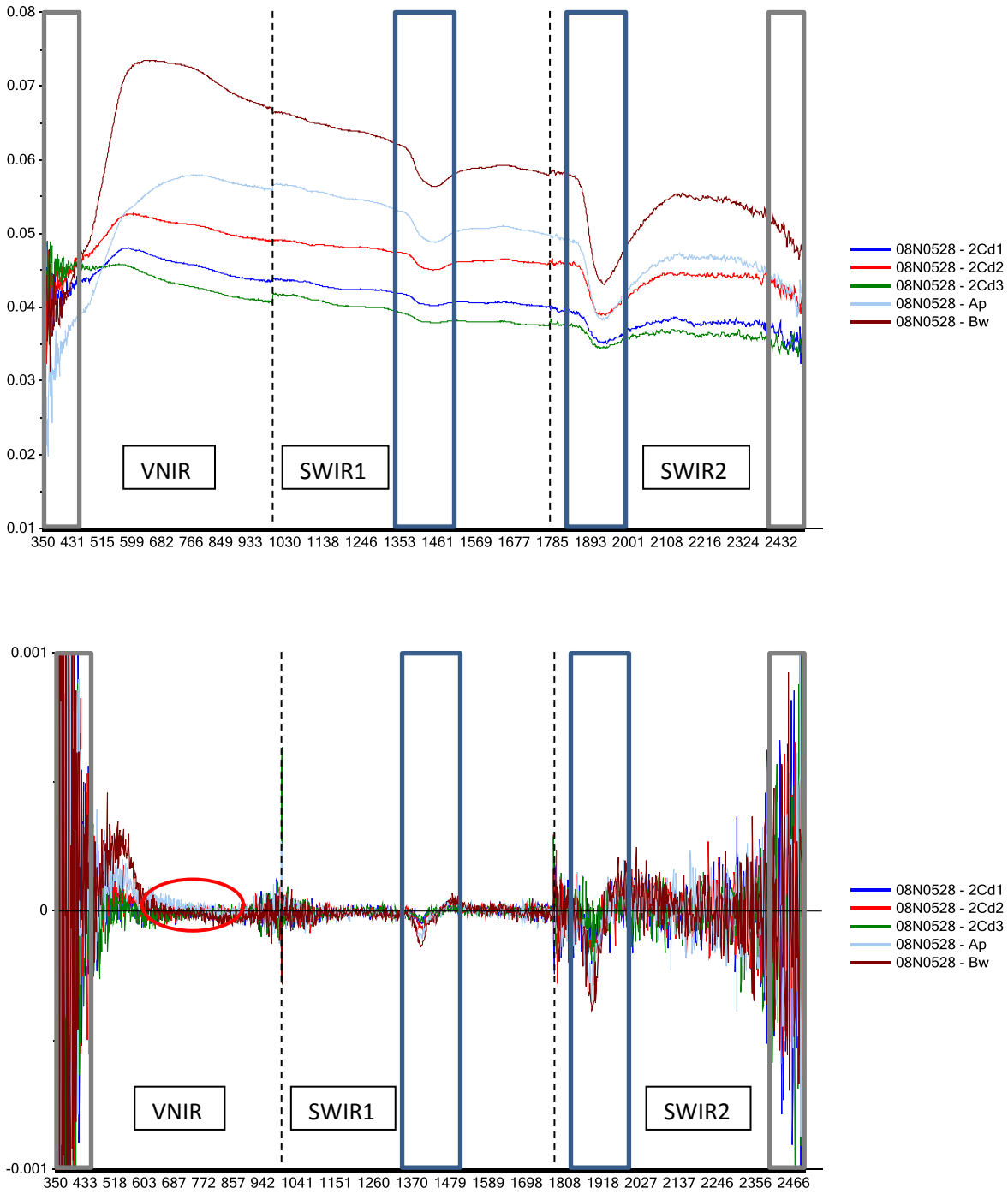


Figure 2.3 – Sample spectra reflectance and first-derivative (Newport soil)

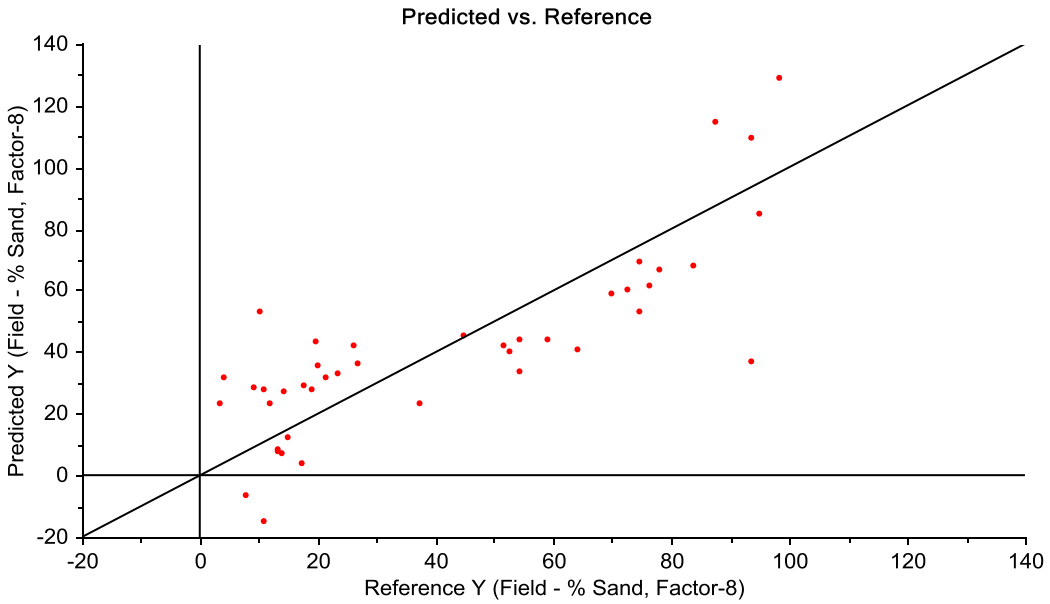
Soil characterization PLS1 models

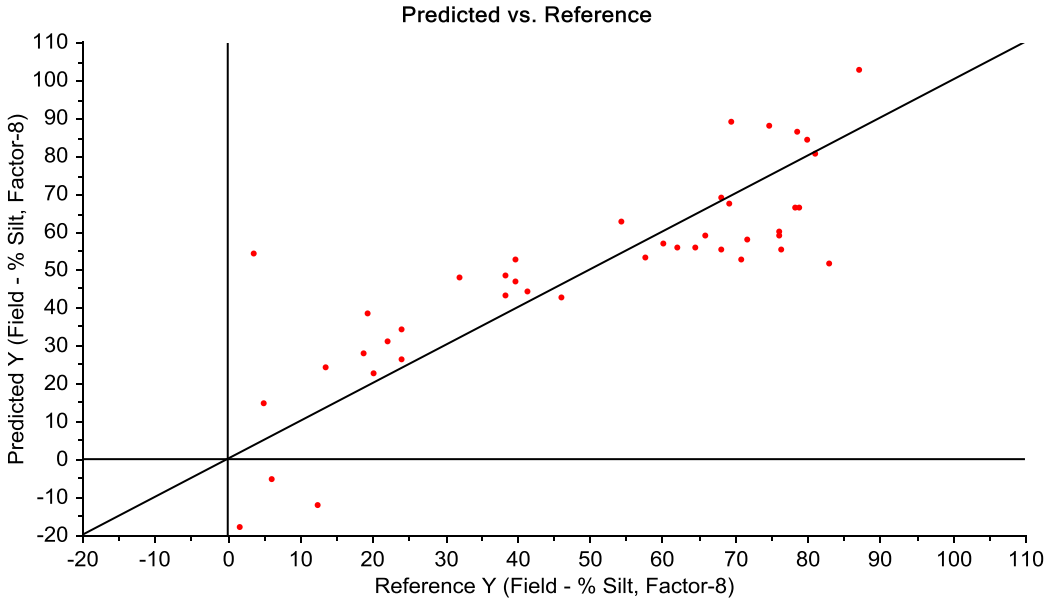
Statistical analyses conducted for selected soil physical and chemical properties in the following tables and plots. Validation PLS1 regression values are presented by spectral pre-treatment (reflectance and first-derivative) (Tables 2.3 and 2.4, respectively). Included are the root-mean-squared errors of prediction (RMSEP), the trendline coefficients in the form $y = Slope \times x + Offset$ and correlation values. In addition, plots showing the measured versus the predicted soil properties values are included for the reflectance and first-derivative data (Figures 2.4 and 2.5, respectively).

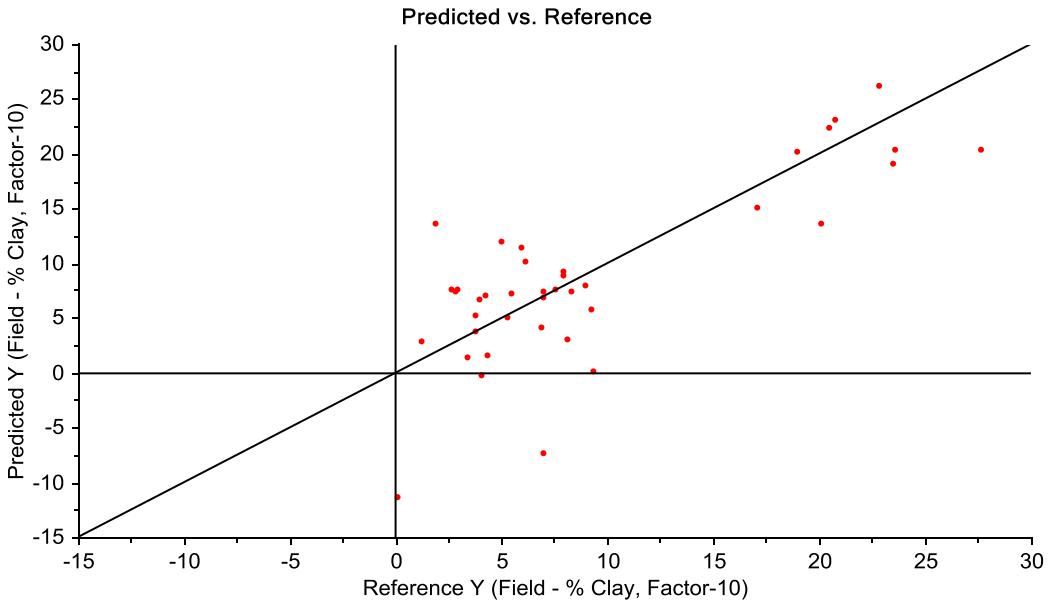
Reflectance

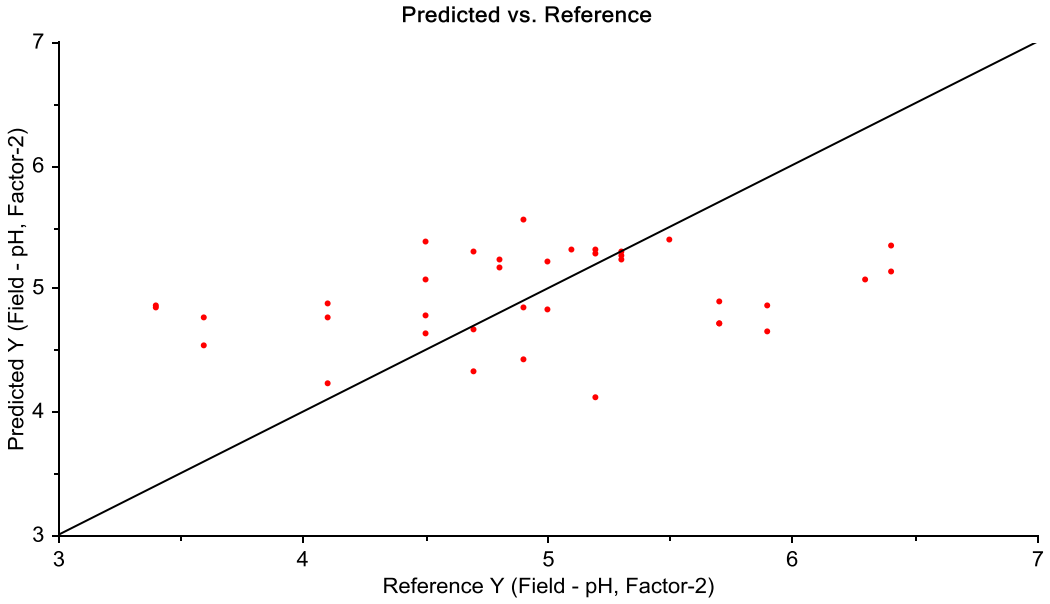
Property	RMSEP	Slope	Offset	Correlation
% Sand	18.89	0.77	9.39	0.80
% Silt	14.83	0.80	9.78	0.84
% Clay	5.00	0.85	0.99	0.79
pH	0.75	0.13	4.32	0.28
CEC	4.95	0.55	3.10	0.71
Activity	1.84	0.34	0.98	0.48
% C	11.84	0.35	6.12	0.50

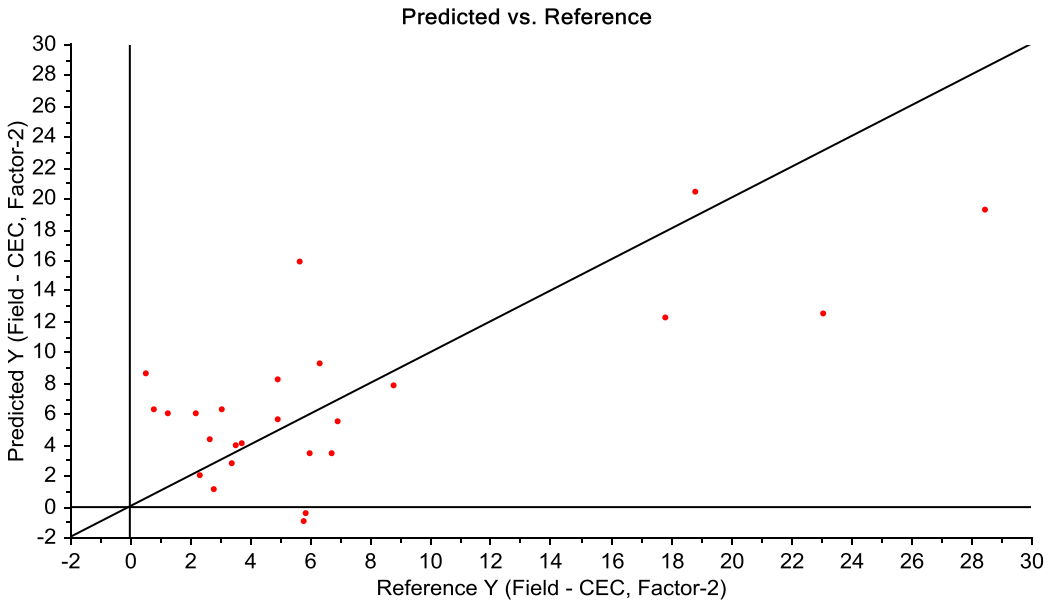
Table 2.3 – Prediction diagnostics for soil characterization PLS1 models (reflectance)

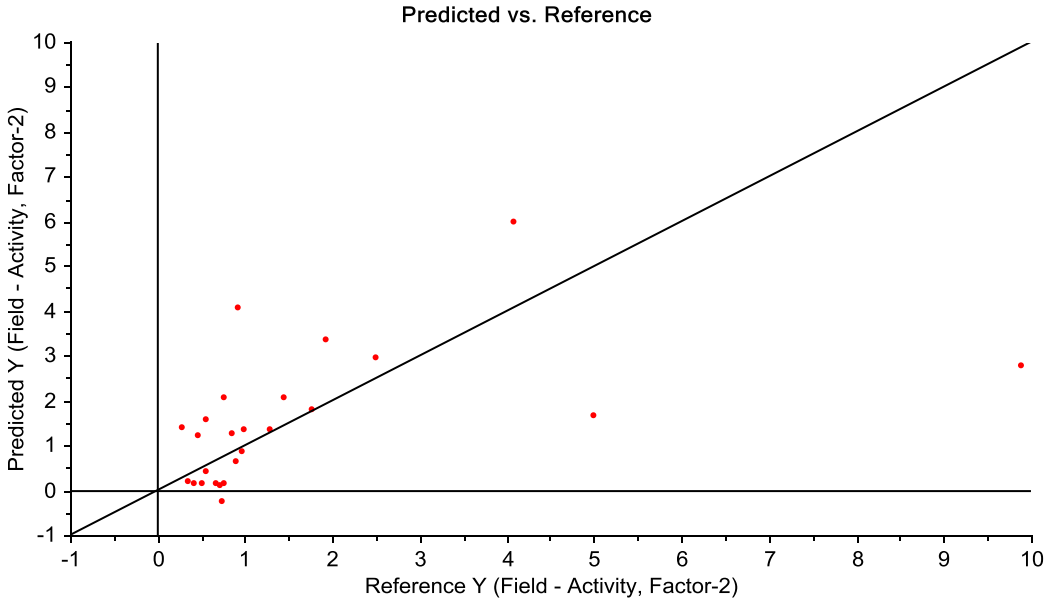

Predicted vs. Reference
Predicted Y (Field - % Sand, Factor-8)
Reference Y (Field - % Sand, Factor-8)
140
120
100
80
60
40
20
0
-20
-20
0
20
40
60
80
100
120
140


Predicted vs. Reference
Predicted Y (Field - % Silt, Factor-8)
Reference Y (Field - % Silt, Factor-8)
110
100
90
80
70
60
50
40
30
20
10
0
-10
-20
-20
-10
0
10
20
30
40
50
60
70
80
90
100
110


Predicted vs. Reference
Predicted Y (Field - % Clay, Factor-10)
Reference Y (Field - % Clay, Factor-10)
30
25
20
15
10
5
0
-5
-10
-15
-15
-10
-5
0
5
10
15
20
25
30


Predicted vs. Reference
Predicted Y (Field - pH, Factor-2)
Reference Y (Field - pH, Factor-2)
7
6
5
4
3
3
4
5
6
7


Predicted vs. Reference
Predicted Y (Field - CEC, Factor-2)
Reference Y (Field - CEC, Factor-2)
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-2
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30


Predicted vs. Reference
Predicted Y (Field - Activity, Factor-2)
Reference Y (Field - Activity, Factor-2)
10
9
8
7
6
5
4
3
2
1
0
-1
-1
0
1
2
3
4
5
6
7
8
9
10

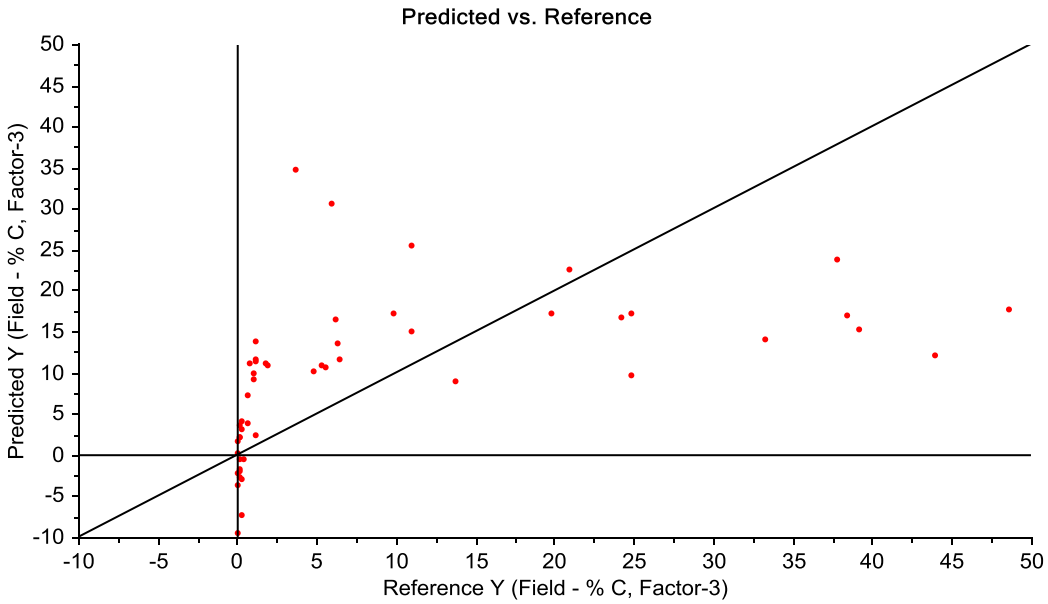
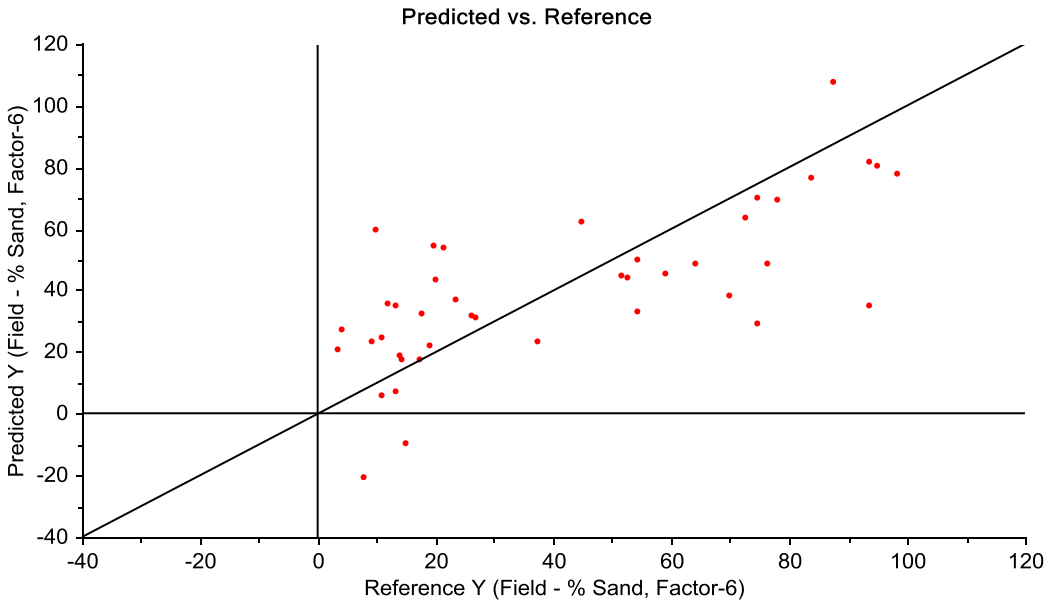


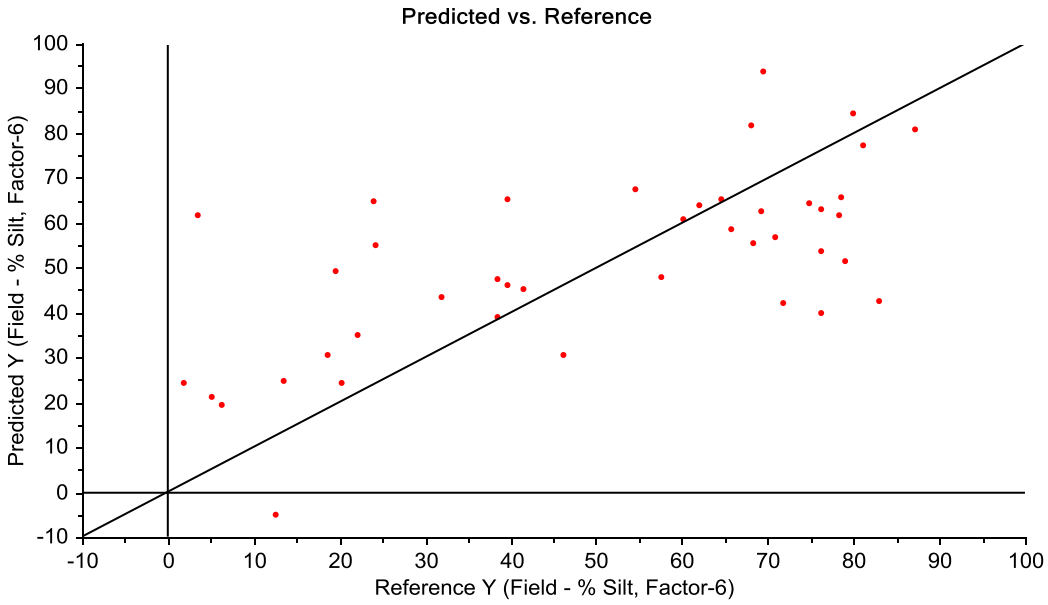
Figure 2.4 – Predicted vs. measured for soil characterization properties (reflectance)

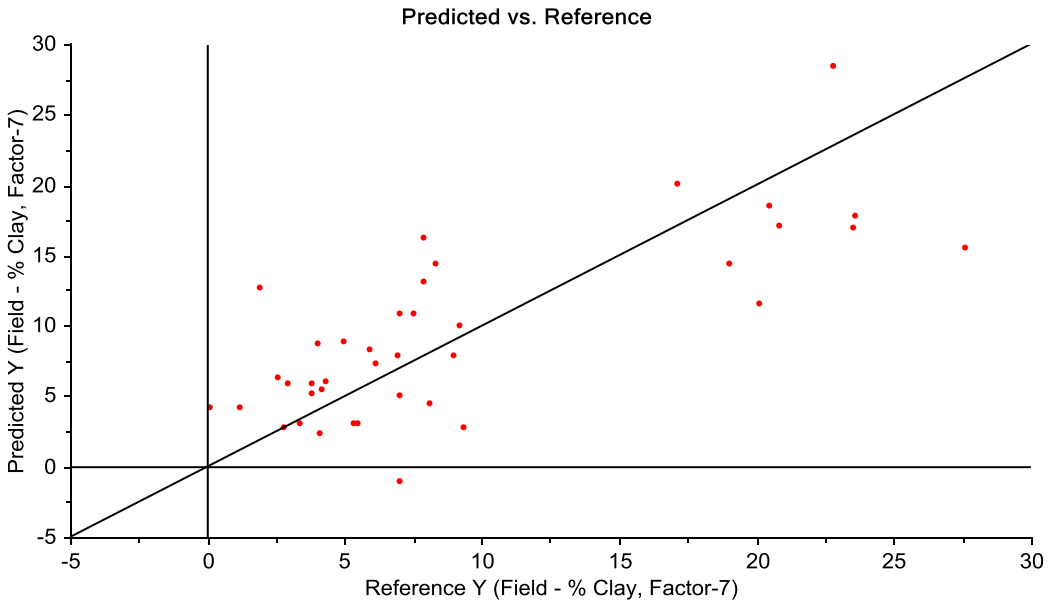
First-derivative

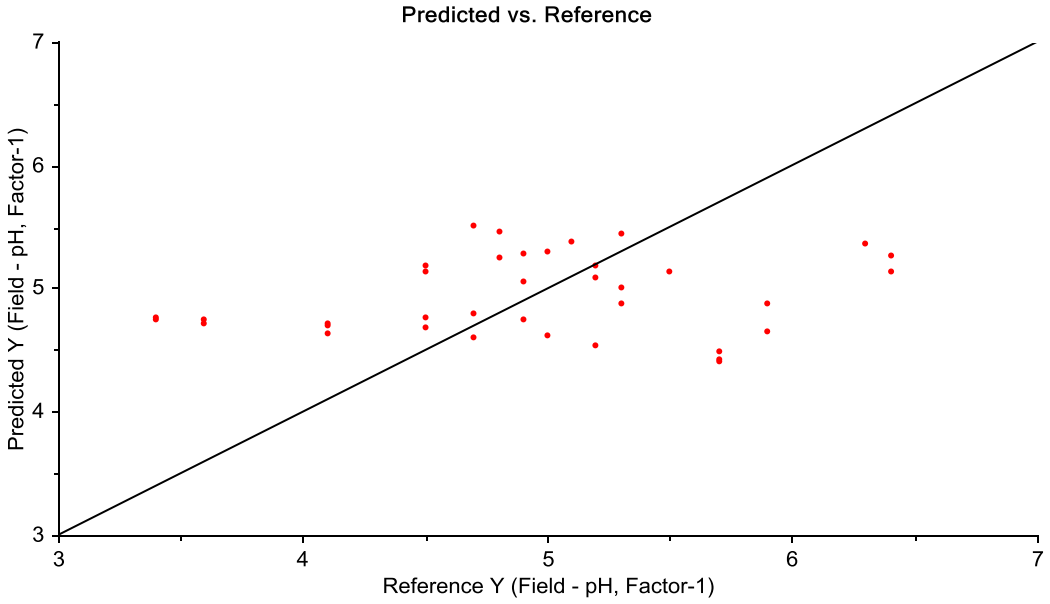
Property	RMSEP	Slope	Offset	Correlation
% Sand	21.62	0.59	16.27	0.72
% Silt	20.21	0.49	27.07	0.66
% Clay	4.80	0.64	3.57	0.76
pH	0.77	0.08	4.53	0.20
CEC	4.74	0.64	2.86	0.75
Activity	1.87	0.25	1.22	0.42
% C	10.52	0.47	4.81	0.63

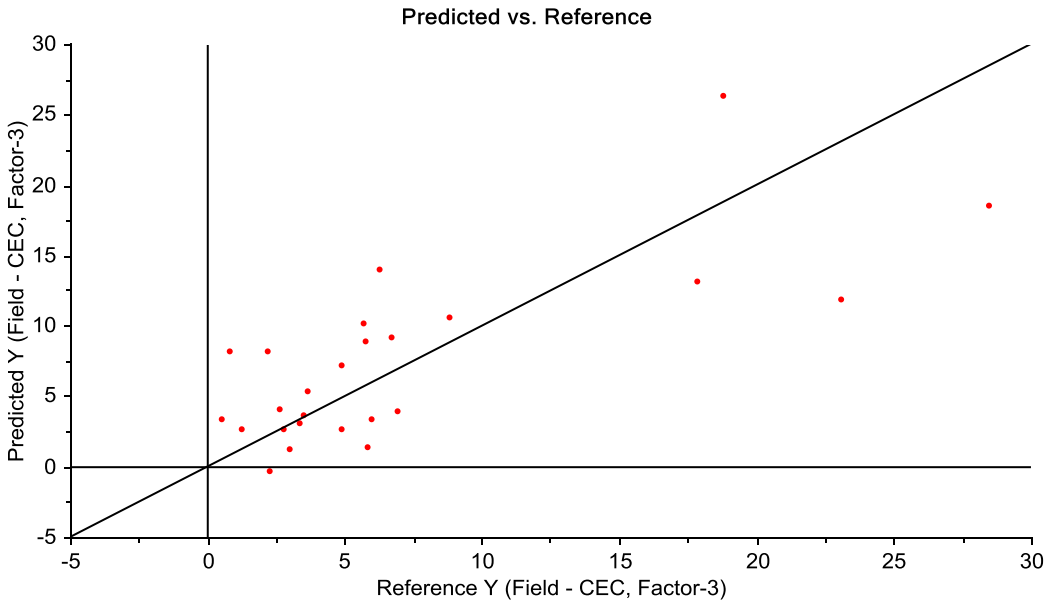
Table 2.4 – Prediction diagnostics for soil characterization PLS1 models (first-derivative)

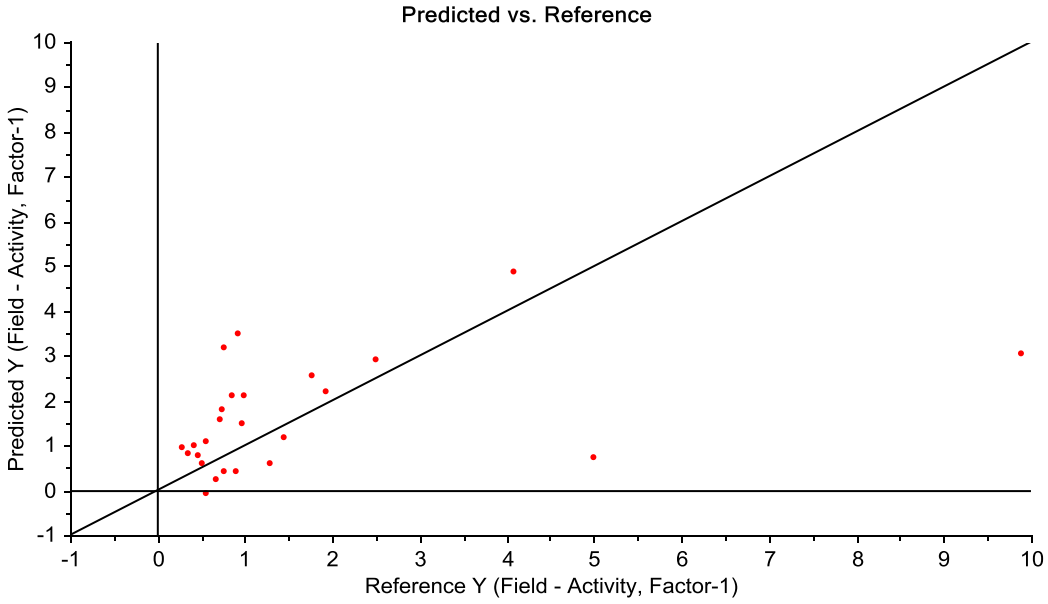

Predicted vs. Reference
Predicted Y (Field - % Sand, Factor-6)
Reference Y (Field - % Sand, Factor-6)
120
100
80
60
40
20
0
-20
-40
-40
-20
0
20
40
60
80
100
120


Predicted vs. Reference
Predicted Y (Field - % Silt, Factor-6)
Reference Y (Field - % Silt, Factor-6)
100
90
80
70
60
50
40
30
20
10
0
-10
-10
0
10
20
30
40
50
60
70
80
90
100


Predicted vs. Reference
Predicted Y (Field - % Clay, Factor-7)
Reference Y (Field - % Clay, Factor-7)
30
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25
30


Predicted vs. Reference
Predicted Y (Field - pH, Factor-1)
Reference Y (Field - pH, Factor-1)
7
6
5
4
3
3
4
5
6
7


Predicted vs. Reference
Predicted Y (Field - CEC, Factor-3)
Reference Y (Field - CEC, Factor-3)
30
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25
30


Predicted vs. Reference
Predicted Y (Field - Activity, Factor-1)
Reference Y (Field - Activity, Factor-1)
10
9
8
7
6
5
4
3
2
1
0
-1
-1
0
1
2
3
4
5
6
7
8
9
10

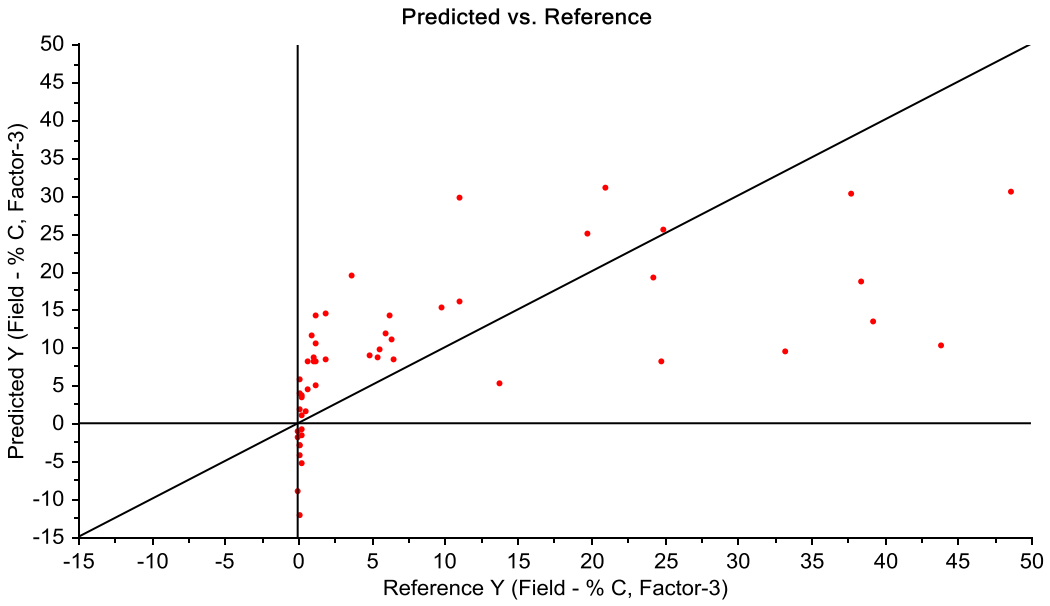


Figure 2.5 – Predicted vs. measured for soil characterization properties (first-derivative)

Interpretations

Since the RMSEP for each property (Tables 2.3 and 2.4) falls within the standard deviation of reference values for the same (Table 2.2), proximal diffuse reflectance spectroscopy can be recommended for use in field soil survey. The standard deviation of % sand is 31.09, while the RMSEP for reflectance models is 18.89 and 21.62 for first-derivative models. The standard deviation of % silt is 26.83, while the RMSEP for reflectance models is 14.83 and 20.21 for first-derivative models. The standard deviation of % clay is 7.41, while the RMSEP for reflectance models is 5.00 and 4.80 for first-derivative models. The standard deviation of pH is 0.82, while the RMSEP for reflectance models is 0.75 and 0.77 for first-derivative models. The standard deviation of CEC is 7.04, while the RMSEP for reflectance models is 4.95 and 4.74 for first-derivative models. The standard deviation of activity is 1.94, while the RMSEP for reflectance models is 1.84 and 1.87 for first-derivative models. The standard deviation of % C is 13.49, while the RMSEP for reflectance models is 11.84 and 10.52 for first-derivative models. Reflectance models performed better for % sand, % silt, pH and activity. First-derivative models performed better for % clay, CEC and % C.

The soils sampled have relatively wide range of sand (95%) and silt (85%) content, contributing to good model performance. The low range of clay in the soil samples (28%) may have degraded model prediction. This also holds for pH, which always has a narrow numeric range and additionally is an indirectly sensed soil property. The remaining soil properties (CEC, activity, % C) were all underpredicted in the PLS1 regression models. Given the low sample size due to the constraints of the soil survey schedule and gaps in soil characterization data, it is highly possible that proximal diffuse reflectance spectroscopy can be suitable for these properties. Additionally, the heterogeneity of the soil samples may account for deficiencies in model characteristics. In particular, the combination of dystrudepts, sulfaquents and sulfihemists may also account for model performance.

Statistically, Figures 2.4 and 2.5 indicate that a relatively low number of principal components (1 – 8) were required to optimize model performance. Additionally, models computed using first-derivative spectra required fewer principal components (1 – 7), than those computed using reflectance spectra (2 – 8). The normalizing effect of the first-derivative may have improved the quality of the predictors, yielding an optimized model at earlier principal components. In terms of point scatter, the plots reflect the underlying skewness and kurtosis (Table 2.2) of the reference data. For example, CEC, activity and % have high levels of skewness (1.99, 3.41 and 1.62 respectively) and kurtosis (3.37, 13.04 and 1.48 respectively), leading to a high number of samples clustering near the prediction axis and underpredicted values for samples with reference values at the top end of their range.

Summary and conclusion

Proximal diffuse reflectance spectroscopy is recommended for use in benchmark soil survey, subject to the limitations described above. Practitioners will need to evaluate published RMSEP values within the context of soil taxonomic requirements (Soil Survey Staff, 2010). Since verbal taxonomic designations have a quantitative basis, errors in prediction could yield substantially different classifications. However,

as reference soil characterization values for the same samples can vary by between major laboratories (Neafsey et al., 2010) errors in classification may be present using conventional techniques.

Given sample heterogeneity, official prediction models should be calculated within soil series. Ideally, all of the spectra used to compute these models should be collected in the field. This will ensure optimal sample property distributions for model development. The primary challenge to implementing this recommendation is the soil survey schedule. Soil survey staff is required to map and sample increasingly larger areas every year, reducing the number of pedons which can be sampled during any given day with the same level of effort. Furthermore, in southern New England, the sampling season is restricted by water saturation late into the spring and by relatively early freezes. Accordingly, the development of spectral prediction models will be a multi-year, or even decadal, effort or will require an increase in the intensity of soil survey operations. Alternatively, exploring ways to exploit NRCS' archived soils could boost sample size and potentially reduce initial investment costs and minimize disruption to existing surveying schedules. Such efforts would need to address the differences between spectra acquired for field-moist and air-dried soil samples, as well as the potential of sample degradation while in storage.

These recommendations will require a substantial, though feasible investment in terms of human effort, capital and institutional policies. While a detailed costing is beyond the scope of this study, proximal diffuse reflectance spectroscopy has the potential to reduce the cost of deriving soil characterization data and information and support soil survey in the field. Additionally, the future role of mid-infrared (MIR) spectroscopy in the field cannot be discounted (Viscarra Rossel et al., 2006). MIR is typically laboratory-based and covers the 2500 – 20 000 nm waveband. Its excellent performance for numerous properties, including carbon fractions, continues to make it appealing. Proximal diffuse reflectance spectroscopy has become more portable and less expensive; it is possible that MIR may follow a similar path.

References

ASD Inc. 2012. What is a derivative spectrum? http://www.asdi.com/resource-center/faqs/what-is-a-derivative-spectrum (accessed 28 Jan. 2012).

Giasson, E., C. van Es, A. van Wambeke and R.B. Bryant. 2000. Assessing the economic value of soil information using decision analysis techniques. Soil Sci. 165:971.

National Cooperative Soil Survey. 2012. National cooperative soil characterization database. http://ssldata.nrcs.usda.gov/ (accessed 28 Jan. 2012).

Neafsey, E.J. 2008. Hyperspectral Sensing of Soil Pedons for Soil Classification and Survey. M.S. thesis, Cornell Univ., Ithaca.

Neafsey, E.J., S.D. DeGloria, M.W. Havens, W.D. Philpot and P.J. Sullivan. 2010. Comparing Physical and Chemical Soil Properties Measurement: CNAL vs. NSSC. Soil Survey Horizons 51: 22-23.

Soil Survey Staff. 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.

U.S. Department of Agriculture, Natural Resources Conservation Service. 2012. National Soil Survey Handbook, title 430-VI. http://soils.usda.gov/technical/handbook/ (accessed 28 Jan. 2012).

USDA. 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S.D.A Agriculture Handbook #296. Natural Resources Conservation Service. United States Department of Agriculture. 669p. [http://soils.usda.gov/survey/geography/mlra, last accessed 04Jan12]

Vasques, G.M. and J.O. Sickman. 2009. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. Soil Sci. Soc. Am. J. 73:176.

Vasques, G.M., S. Grunwald and J.O. Sickman. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146:14-25.

Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics Intellig. Lab. Syst. 90:72-83.

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75.

Viscarra Rossel, R.A., Y.S. Jeon, I.O.A. Odeh and A.B. McBratney. 2008. Using a legacy soil sample to develop a mid-IR spectral library. Aust. J. Soil Res. 46:1-16.

Walvoort, D.J.J. and A.B. McBratney. 2001. Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. p. 503-507. *In* Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. ECPA, 2001.

Chapter 3: Characterizing soil properties using diffuse reflectance spectroscopy for subaqueous soil survey in coastal southern New England

Abstract

Subaqueous soils develop in marine, lacustrine or riverine environments (to 2.5 m, the maximum rooting water depth for submerged aquatic vegetation) and play a crucial role in the ecosystems within which they developed. Proximal visible-near infrared diffuse reflectance spectroscopy (VNIR DRS) offers rapid, accurate and nondestructive estimation of key soil properties. Very few studies have used this technology for the characterization of subaqueous soils. Specific study objectives are to: (1) develop methods to apply DRS to core samples of subaqueous soils; (2) estimate selected properties of subaqueous soils (% sand, total sulfide, soil organic carbon (% SOC), salinity and pH); and (3) make recommendations on the integration of DRS in subaqueous soil survey procedures. This research seeks to test the utility of diffuse-reflectance visible-near infrared diffuse reflectance spectroscopy (VNIR DRS) for the development of accurate soil property prediction models (using PLS1 regression) for subaqueous soils in coastal southern New England. PLS1 regression models computed using reflectance spectra had RMSEP values of 25.81 for % sand, 1355.72 for total sulfide, 22.94 for % SOC, 15.06 for salinity and 0.76 for pH; those using first-derivative spectra had RMSEP values of 26.01 for % sand, 1317.72 for total sulfide, 20.63 for % SOC, 12.17 for salinity and 0.72 for pH. Proximal diffuse reflectance spectroscopy is recommended for use with subaqueous soil survey procedures.

Introduction

Subaqueous soils develop in marine, lacustrine or riverine environments and play a crucial role in the ecosystems within which they developed. Proximal visible-near infrared diffuse reflectance spectroscopy's (VNIR DRS) offers rapid, accurate and nondestructive estimation of key soil properties. Very few studies have used this technology for the characterization of subaqueous soils. Specific study objectives are to: (1) develop methods to apply DRS to core samples of subaqueous soils; (2) estimate selected properties of subaqueous soils (% sand, total sulfide, soil organic carbon (% SOC), salinity and pH); and (3) make recommendations on the integration of DRS in subaqueous soil survey procedures.

Ecological significance

Given their important positions on the landscape, formalizing our understanding and mapping of subaqueous soils will aid in the management of coastal resources (Bradley and Stolt, 2002; Fonseca et al., 1998; Desbonnet et al., 1999; Koch, 2001). The soil science community is working with the marine, lacustrine and riverine ecology communities to understand the relationships between species dynamics and the soils their ecosystems depend upon. Expanding soil mapping to these areas have the potential to aid in the site-appropriate restoration of submerged aquatic vegetation and commercially important species such as clams, oyster and scallops (Demas and Rabenhorst, 1999).

Economic importance

Human activity has historically and will likely continue to be concentrated in coastal areas. It is important for society to understand the engineering characteristics of subaqueous soils, including their viability for serving as a base for transportation infrastructure and buildings, as well as their potential uses as construction materials (Bradley and Stolt, 2002). "Subaqueous soil resource inventories could also be significant to identifying estuarine areas with hazards associated with the [terrestrial] disposal of dredge materials with a high potential for acid-sulfate weathering" (Demas and Rabenhorst, 1999).These materials, calculated as total sulfides include acid volatile sulfides (AVS), which take the

forms of 'Amorphous FeS,' Mackinawite (≅ FeS) and Greigite (≅ Fe_3S_4) and chromium reducible sulfur (CRS).

Soil survey expansion

Previous research has resulted in the expansion of the definition of soils to include subaqueous soils (Demas and Rabenhorst, 1999). The *Keys to Soil Taxonomy* has expanded the definition of soils to include "shallow water" to a depth of 2.5 m (Soil Survey Staff, 2010).

Proximal diffuse reflectance spectroscopy offers the promise of rapid, accurate and nondestructive estimation of key soil properties. The use of this technology for the characterization of subaqueous soils is not sufficiently documented in the literature. Viscarra Rossel et al. (2006) and Walvoort and McBratney (2001) provide excellent historical treatment for the use of proximal diffuse reflectance spectroscopy in terrestrial soil characterization. A major outstanding issue in spectral quality is the saturated moisture levels of these soils (Lobell and Asner, 2002). The goal of this study is to test the utility of visible-near infrared diffuse reflectance spectroscopy (VNIR DRS) for the development of accurate soil property prediction models for subaqueous soils in coastal southern New England. Specific objectives are to: (1) develop methods to apply DRS to core samples of subaqueous soils; (2) estimate selected properties of subaqueous soils (% sand, total sulfides, soil organic carbon salinity and pH); and (3) make recommendations on the integration of DRS in subaqueous soil survey procedures.

Materials and methods

Study area

The project study area is located within the Northeastern Forage and Forest Land Resource Region. Topography is characterized by plateaus, plains and mountains. The climate is cool and humid, experiencing an average precipitation of 865 – 1575 mm per year. Average annual temperature is 4 – 9 °C. There are 130 – 200 freeze days, with 110 in the highest elevations and 240 in some coastal areas.

Total withdrawals of freshwater from the region are, on average, 72 ML/d, 93 % of which is from surface water and 7 % from groundwater. Timber and manufacturing interests consume approximately 78 % of this total, with the rest for public supply.

Soils in this region are primarily Entisols or Spodosols, commonly with a fragipan. In areas with limey geology, Alfisols occur with a fragipan. Higher elevations are characterized by Ochrept and Orthods, while lower regions feature Aqualfs, Aquepts and Histosols. Fluvents in floodplains are limited in extent but are locally important. The temperature regimes are frigid or mesic, udic in terms of soil moisture and exhibit a mixed mineralogy.

The primary land cover is forest (especially in areas of rugged relief), with 98 % of the land in private ownership. Logging, for timber and pulpwood, is a commercially important activity, as is the production of Christmas trees and maple syrup. The region grows forage and grains to feed local dairy stock, however local conditions can allow for the production of food crops for human consumption. Wildlife habitat and recreation are major land uses, especially where stoniness and steep slope has precluded other uses (USDA, 2006).

DRS measurements

To ensure data quality a list was compiled of subaqueous soil cores which have descriptions associated with them. Soil cores are stored in the halved 75 - 100 mm vibracore in which they were initially acquired. To preserve atmospheric conditions, these vibracores are wrapped tightly in foodservice grade plastic film. The ends of the film are secured using clear or duct tape. As an archival measure, the pedon ID is written in permanent marker on the plastic film. The wrapped soil cores are stored in 1750 mm plastic trays, the ends of which is secured using foam padded caps and electrical tape. Again, the pedon ID and collection location was noted on the caps.

Subaqueous soil core spectra were measured with an ASD FieldSpec Pro FR spectrometer (Analytical Spectral Devices, Inc., Boulder, CO[5], using a contact probe directly against the soil profile. The spectrometer collects data at a spectral resolution of 1 nm. It is also a three-spectrometer system, with detectors for the VNIR (350 – 1000 nm), SWIR1 (1001 – 1780 nm) and SWIR2 (1781 – 2500 nm) spectral regions. Discontinuities can be observed (Figure 3.1) in the spectra in transitioning between these regions.

A clean washcloth was used to wipe away any residue left on the contact probe lens after each sample was collected. Water evaporation was kept to a minimum through rapid sampling and keeping the light source off between horizon sampling to minimize the temperature of the soil sample and contact probe, respectively. The laptop computer was connected to the spectrometer in the same manner as in the field. A white reference standard (spectralon) was used to optimize the instrument before each session and to recalibrate it after samples from each soil pedon's horizons had been scanned. In addition, only one set of five spectra (each an instantaneous average of ten spectra) is taken for each horizon, due to the width of the soil cores.

Sampling for laboratory analysis

After profile description and spectral acquisition were complete, the Natural Resources Conservation Service (NRCS) soil survey team collected soil samples for the pedon for submission to the National Soil Survey Laboratory (NSSL) at the National Soil Survey Center (NSSC) in Lincoln, NE. Soil samples were analyzed by the NSSL and the NRCS-RI/University of Rhode Island (URI) to determine their physical, chemical, and mineralogical properties according to NRCS and URI procedures (U. S. Department of Agriculture, Natural Resources Conservation Service, 2012; National Cooperative Soil Survey, 2012).

[5] <http://www.asdi.com>

Data analysis

To facilitate data analysis, the instrument native files (spectral reflectance calculated by the ASD software) for each horizon were imported into The Unscrambler software program,[6] a multivariate statistics package which specializes in spectral processing and analysis. Reflectance spectra can also be imported to universally-readable ASCII files by ASD's ViewSpecPro, but direct importation provides a more efficient workflow and ensures the preservation of variable names.

The resulting files for each horizon were averaged (across the five spectra collected for each horizon), yielding one spectrum for each horizon. Averaging accomplishes instrument noise reduction and accounts for variance in the samples.

First derivatives were calculated using the difference between the reflectance values associated with adjoining wavelengths at a resolution of 1 nm to preserve full spectral information. As soil moisture can reduce the amplitude of soil spectra (particularly an issue in subaqueous soil samples), derivative analysis is used to normalize these amplitudes, highlighting the constituents of interest (ASD Inc., 2012).

PLS1 regression modeling

Partial least squares regression was used to evaluate the statistical contributions of many X-variables (predictors) to one Y-variable (responses). In this study, the reflectance values for the waveband 350 – 2500 nm comprised the set of X-variables. Excluded due to falloff in detector sensitivity were the wavebands 350 – 450 nm and 2400 – 2500 nm (marked in gray in Figure 3.1). Excluded also were water absorption wavebands 1350 – 1500 nm and 1850 – 2000 nm (marked in blue in Figure 3.1). Selected physical and chemical soils properties were the Y-variables. A common statistical tool in soil chemometrics, PLS1 regression allows the user to determine which spectral wavelengths are correlated with properties of interest (Vasques et al., 2008; Vasques and Sickman, 2009; Viscarra Rossel, 2008).

---

[6] <http://www.camo.com/rt/Products/Unscrambler/unscrambler.html>

Results and discussion

Sample descriptions

Full data descriptions of subaqueous soil samples are located online at Mapcoast[7] and Nesoil.com[8], summarized here for selected soil properties. Based on the median (Table 3.1), the soils in this study are characterized by high sand content. They have moderate total sulfides and low % SOC. They are also slightly to moderately saline as well as slightly alkaline.

There are several features of particular interest in the sample spectral set (Figure 3.1) for a single saturated pedon representing the Nagunt series, reflectance and first-derivative. While not a strong relationship, the primary variable controlling spectral reflectance is sand content. The horizon with the highest maximum reflectance is the Cg4, with 97 % sand, while the lowest reflectance appears in the Cg3 horizon, with a relatively low 82 % sand. These also include the shape of spectra in visible, the slope of spectra in the 2300 – 2500 nm range, the slope differences at the wings of the water absorption bands (boxed in blue, Cyan and blue curves), the depth of water absorption bands and the overall slope of spectra from 700 – 1300 nm. For derivative spectra, note that the two spikes (~1000 nm, 1750 nm) are transitions between spectrometers in the ASD. These transitions are apparent as calibration errors in the reflectance spectra in the top portion of the figure.

[7] <http://www.ci.uri.edu/projects/mapcoast/data.html>
[8] <http://nesoil.com/sas/index.htm>

	Subaqueous - % Sand	Subaqueous - Total Sulfide	Subaqueous - % SOC	Subaqueous - Salinity	Subaqueous - pH
Units	%	µg/g	%	dS m-1	pH
#	209	19	179	194	231
Mean	67.90	1717.16	6.25	17.70	7.33
Max	100.86	5853.66	191.86	72.40	8.86
Min	2.80	33.31	0.05	0.12	3.40
Range	98.06	5820.35	191.81	72.28	5.46
Std Deviation	30.83	1751.91	23.49	16.60	1.00
Variance	950.45	3069189.00	551.62	275.60	1.01
RMS	74.54	2418.12	24.24	24.24	7.40
Skewness	-0.75	0.99	6.43	0.86	-1.45
Kurtosis	-0.88	-0.03	45.10	-0.28	2.01
Median	80.84	907.78	0.77	7.87	7.63
Q1	43.83	366.28	0.38	4.33	6.90
Q3	94.00	3163.15	2.56	30.53	8.00

Table 3.1 – Soil characterization properties summary statistics

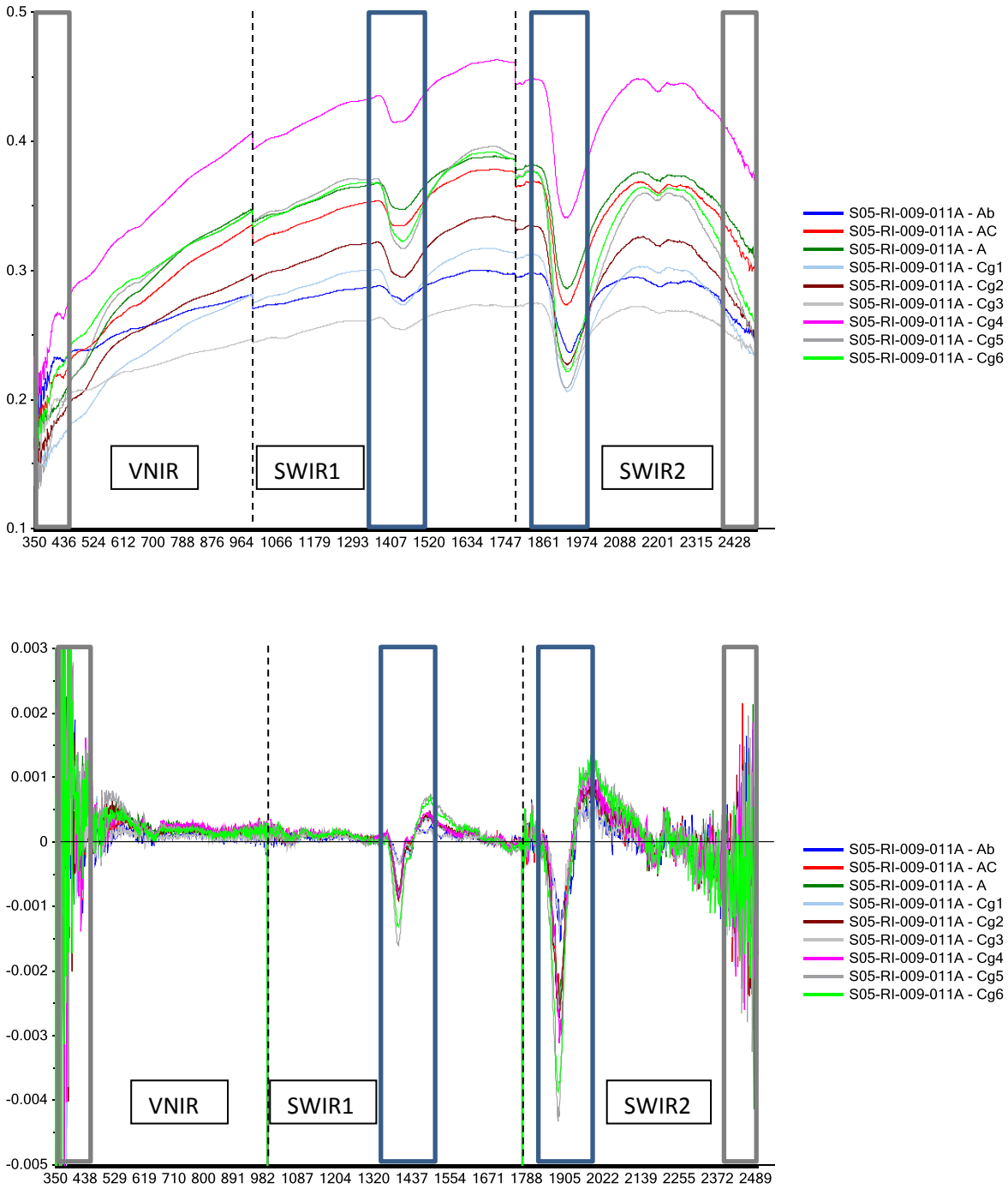


Figure 3.1 – Sample spectra reflectance and first-derivative (Nagunt soil)

Soil characterization PLS1 models

Statistical analyses conducted for selected soil physical and chemical properties in the following tables

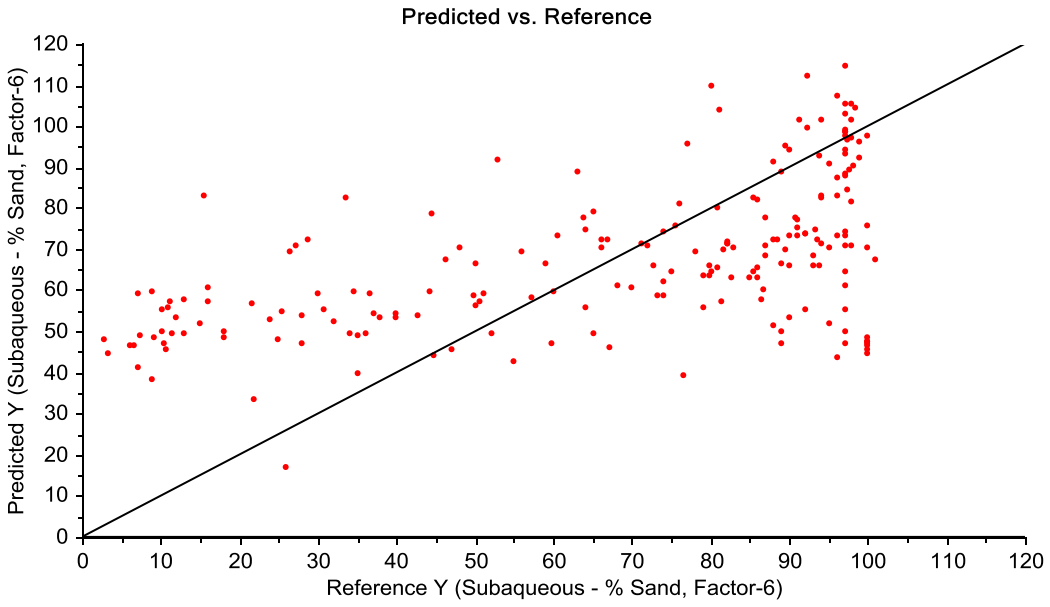
and plots. Validation PLS1 regression values are presented by spectral pre-treatment (reflectance and

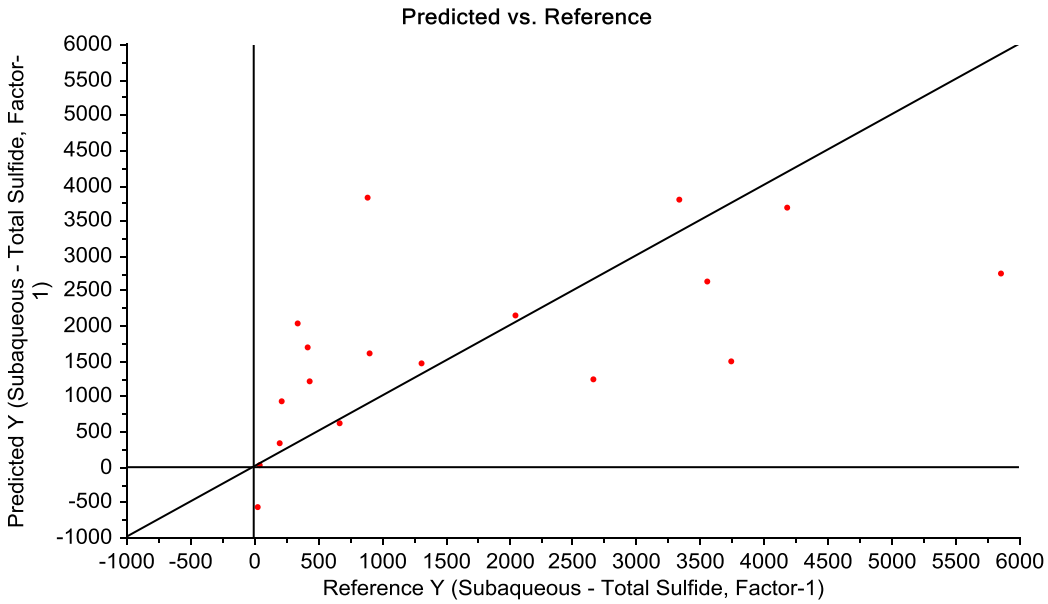
first-derivative) (Tables 3.2 and 3.3, respectively). Included are the root-mean-squared errors of prediction (RMSEP), the trendline coefficients in the form $y = Slope \times x + Offset$and correlation values. In addition, plots showing the measured versus the predicted soil properties values are included for the reflectance and first-derivative data (Figures 3.2 and 3.3, respectively).

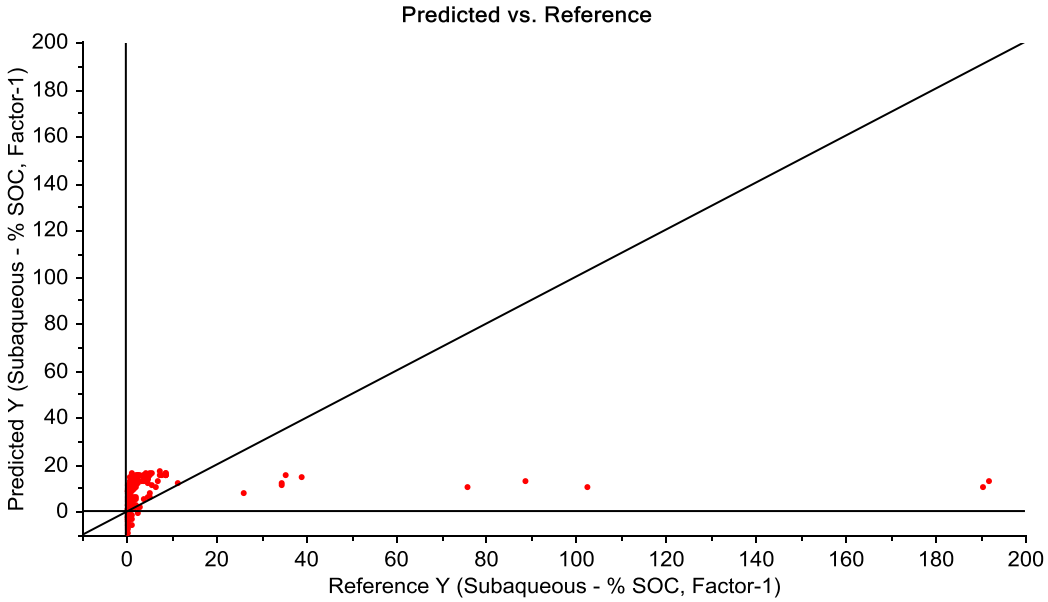
Reflectance

Property	RMSEP	Slope	Offset	Correlation
% Sand	25.81	0.32	46.15	0.55
Total Sulfide	1355.72	0.44	951.93	0.61
% SOC	22.94	0.06	5.84	0.21
Salinity	15.06	0.36	10.96	0.49
pH	0.76	0.51	3.59	0.66

Table 3.2 – Prediction diagnostics for soil characterization PLS1 models (reflectance)




Predicted vs. Reference
Predicted Y (Subaqueous - Total Sulfide, Factor-1)
Reference Y (Subaqueous - Total Sulfide, Factor-1)
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
-1000
-1000 -500 0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000


Predicted vs. Reference
Predicted Y (Subaqueous - % SOC, Factor-1)
Reference Y (Subaqueous - % SOC, Factor-1)
200
180
160
140
120
100
80
60
40
20
0
0 20 40 60 80 100 120 140 160 180 200

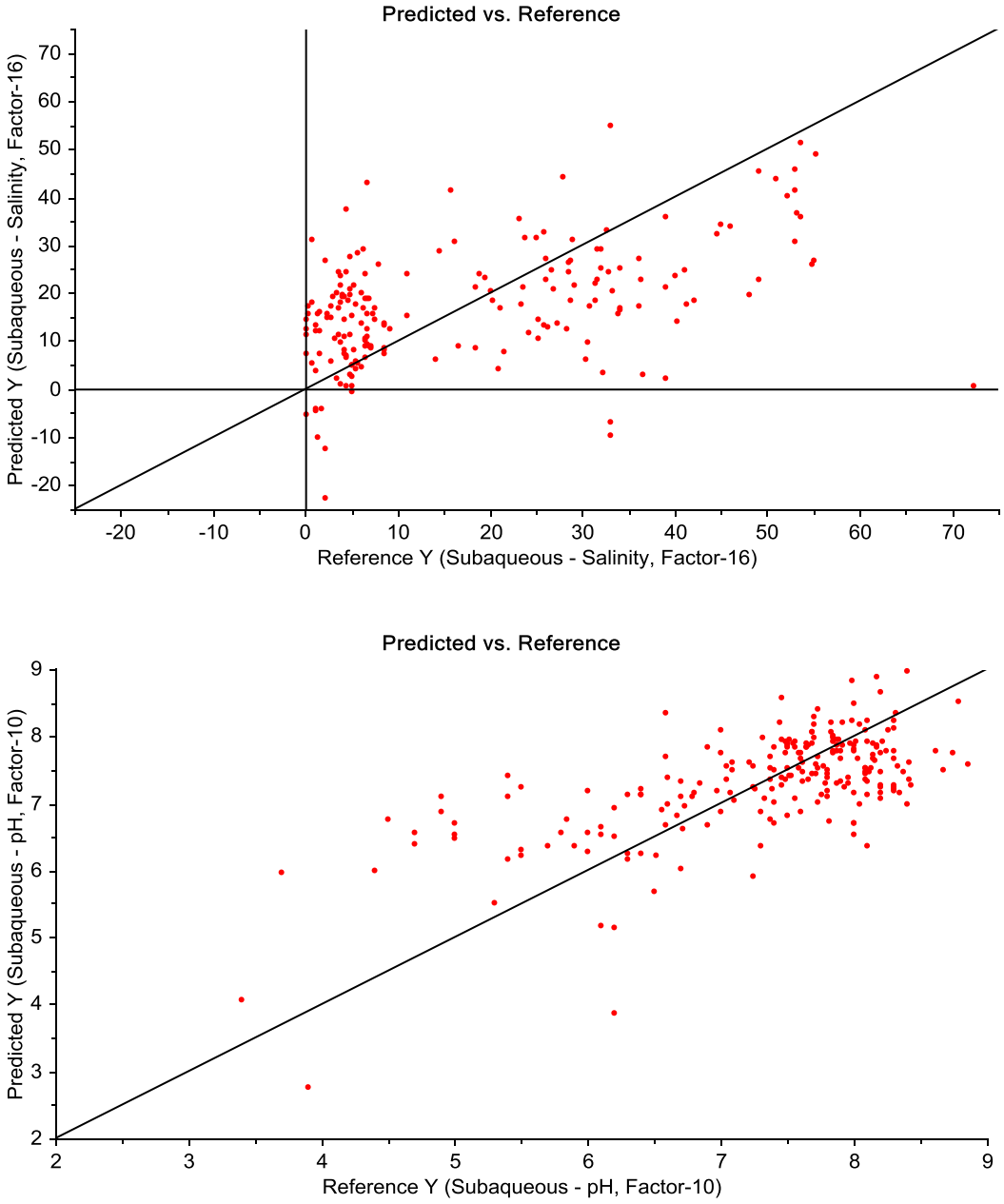
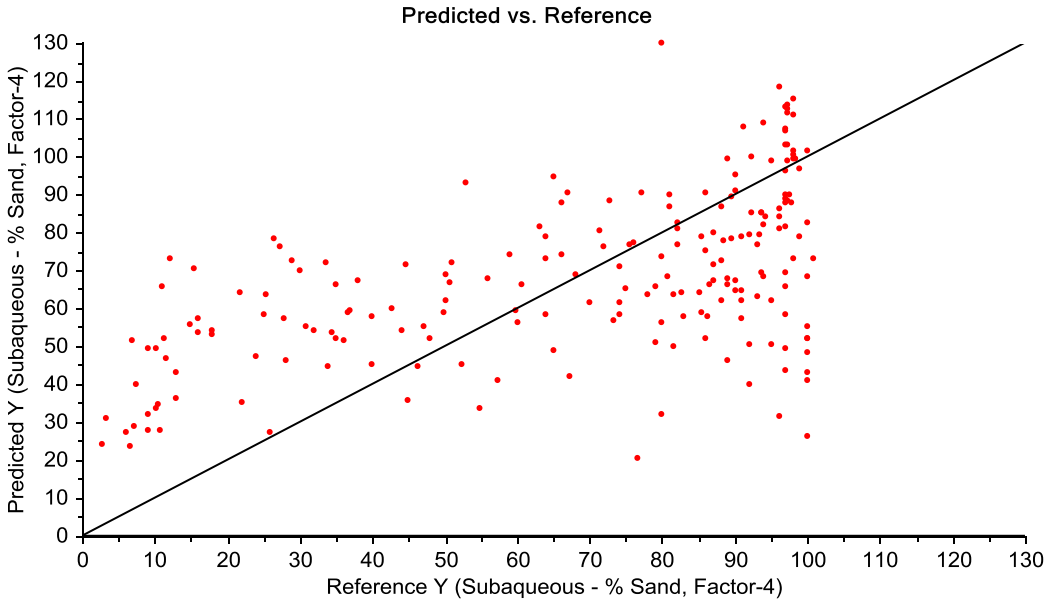


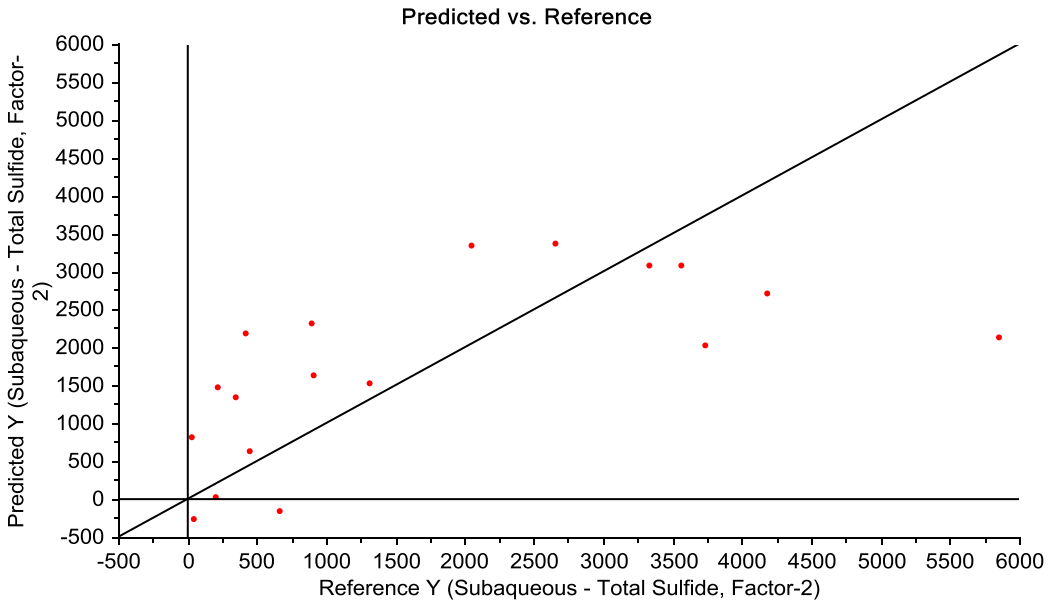
Figure 3.2 – Predicted vs. measured for soil characterization properties (reflectance)

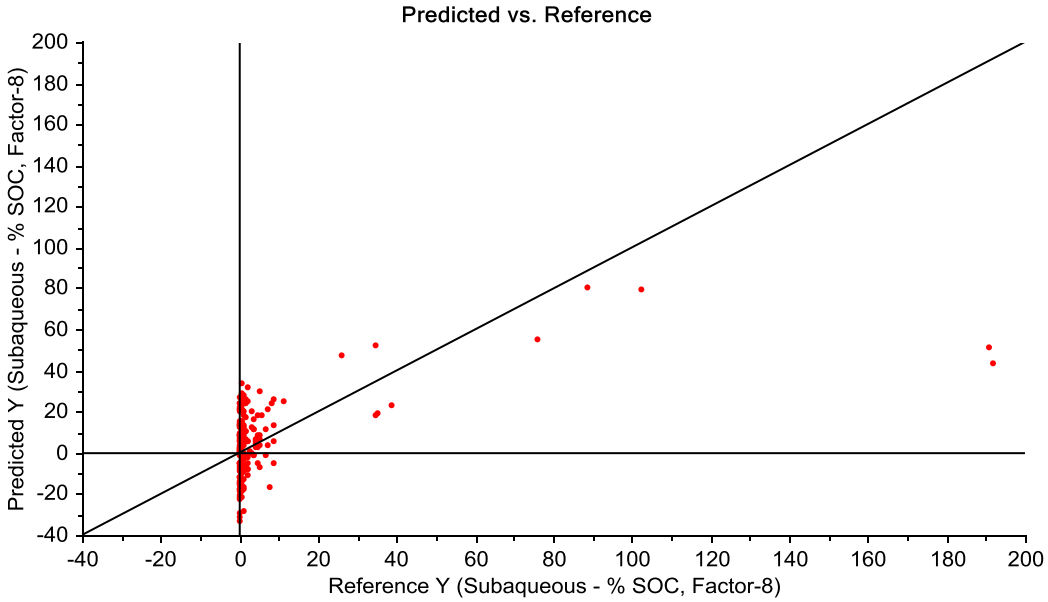
First-derivative

Property	RMSEP	Slope	Offset	Correlation
% Sand	26.01	0.39	41.32	0.56
Total Sulfide	1317.72	0.42	1012.67	0.63
% SOC	20.63	0.39	3.49	0.52
Salinity	12.17	0.51	8.69	0.68
pH	0.72	0.57	3.13	0.71

Table 3.3 – Prediction diagnostics for soil characterization PLS1 models (first-derivative)




Predicted vs. Reference
Predicted Y (Subaqueous - Total Sulfide, Factor-2)
Reference Y (Subaqueous - Total Sulfide, Factor-2)
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
-500
-500
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000


Predicted vs. Reference
Predicted Y (Subaqueous - % SOC, Factor-8)
Reference Y (Subaqueous - % SOC, Factor-8)
200
180
160
140
120
100
80
60
40
20
0
-20
-40
-40
-20
0
20
40
60
80
100
120
140
160
180
200

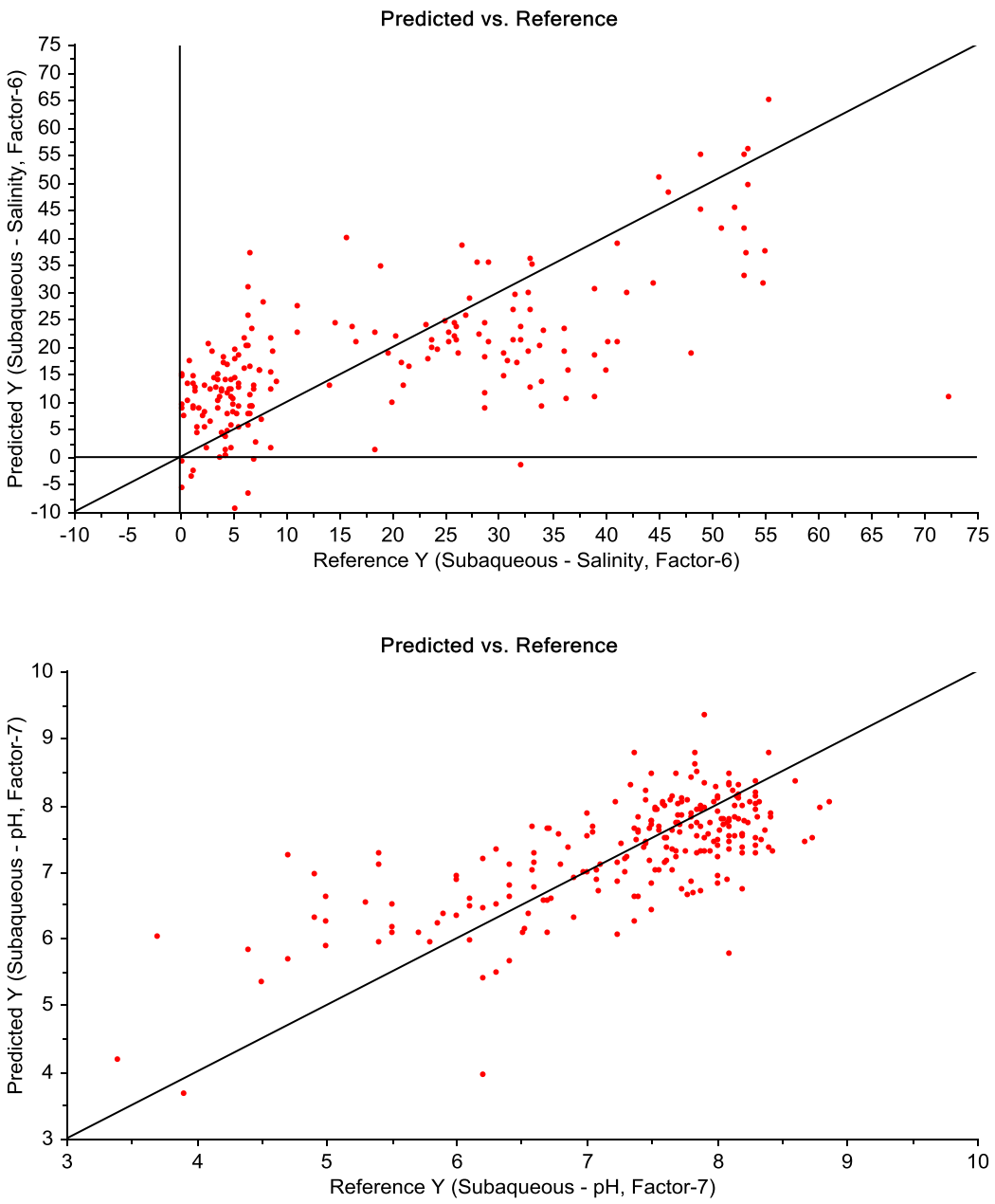


Figure 3.3 – Predicted vs. measured for soil characterization properties (first-derivative)

Interpretations

Since the RMSEP for each property (Tables 3.2 and 3.3) falls within the standard deviation of reference values for the same (Table 3.1), proximal diffuse reflectance spectroscopy is recommended for use in subaqueous soil survey. The standard deviation of % sand is 30.83, while the RMSEP for reflectance models is 25.81 and 26.01 for first-derivative models. The standard deviation of total sulfide is 1751.91,

while the RMSEP for reflectance models is 1355.72 and 1317.72 for first-derivative models. The standard deviation of % SOC is 23.49, while the RMSEP for reflectance models is 22.94 and 20.63 for first-derivative models. The standard deviation of salinity is 16.60, while the RMSEP for reflectance models is 15.06 and 12.17 for first-derivative models. The standard deviation of pH is 1.00, while the RMSEP for reflectance models is 0.76 and 0.72 for first-derivative models. Reflectance models performed better for % sand. First-derivative models performed better for total sulfide, % SOC, salinity and pH.

Given the low number of samples used in computation (n=19), it is significant that the reflectance and first-derivative models performed reasonably well for total sulfides. Given the long-term expense of laboratory processing these results indicate additional samples should be analyzed for total sulfides to enhance the quality of updated prediction models. Also of note is that % sand is over predicted for low values and underpredicted for high values. Percent SOC was highly underpredicted while most of the soil samples had low values for this property. Given the possibility of sample degradation and gaps in soil characterization data, it is highly possible that proximal diffuse reflectance spectroscopy can be suitable for the selected properties. Additionally, the heterogeneity of the soil samples may account for deficiencies in model performance.

Statistically, Figures 3.2 and 3.3 indicate that a relatively moderate number of principal components (1 – 16) were required to optimize model performance. Additionally, models computed using first-derivative spectra required fewer principal components (2 – 8), than those computed using reflectance spectra (1 – 16). The normalizing effect of the first-derivative may have improved the quality of the predictors, yielding an optimized model at earlier principal components. In terms of point scatter, the plots reflect the underlying skewness and kurtosis (Table 3.1) of the reference data. For example, % SOC has high levels of skewness (6.43) and kurtosis (45.10), leading to a high number of samples clustering near the prediction axis and underpredicted values for samples with reference values at the top end of their

range. The highest value horizon in terms of total sulfides (5853.66), an Ab from Wickford Harbor in Rhode Island was also underpredicted.

Summary and conclusion

Proximal diffuse reflectance spectroscopy is recommended for use in subaqueous soil survey, subject to the limitations described above. Practitioners will need to evaluate published RMSEP values within the context of soil taxonomic requirements (Soil Survey Staff, 2010). Since verbal taxonomic designations have a quantitative basis, errors in prediction could yield substantially different classifications. However, as reference soil characterization values for the same samples can vary by between major laboratories (Neafsey et al., 2010) errors in classification may be present using conventional techniques.

Perhaps the greatest obstacle to developing soil characterization prediction models is the lack of both quantity and standardization of soil characterization data. The majority of samples was processed by NRCS-RI and URI and was not submitted to the NSSL for full characterization. Furthermore, critical properties such as total sulfides are expensive and labor intensive to determine in the laboratory, limiting the number of samples which can be processed. Addressing this problem will require significant national support for processing subaqueous soil samples, including financing for NSSL characterization where requested.

Additionally, the spectral collection method described in this paper could be improved. Exposing soil samples to the atmosphere, even for a limited period of time, risks oxidation and sample contamination. The collection approach described by Ben-Dor and Chudnovsky (2008) could be adapted in combination with a vacuum tube to more safely sample vibracored soil samples. Given that subaqueous soil vibracores are often opened onshore, field station installation of potentially more accurate MIR (covering the 2500 – 20 000 nm waveband) instrumentation may be able to provide fast results without requiring the portability that is already available for the 350 – 2500 nm waveband. As subaqueous soil

survey is relatively new and uniquely innovative, this is an excellent point in its history to determine if or how to implement proximal diffuse reflectance spectroscopy in the field.

References

ASD Inc. 2012. What is a derivative spectrum? http://www.asdi.com/resource-center/faqs/what-is-a-derivative-spectrum (accessed 28 Jan. 2012).

Ben-Dor, E.H. and D. Chudnovsky. 2008. A novel method of classifying soil profiles in the field using optical means. Soil Sci. Soc. Am. J. 72:1113.

Bradley, M.P. and M.H. Stolt. 2002. Evaluating methods to create a base map for a subaqueous soil inventory. Soil Sci. 167:222.

Demas, G.P. and M.C. Rabenhorst. 1999. Subaqueous soils: Pedogenesis in a submersed environment. Soil Sci. Soc. Am. J. 63:1250-1257.

Desbonnet, A., T. Kapka, F. Presley, R. Duhaime, J. Barrett, J. Klinger, M. Vincent and L.M. Ernst. 1999. Rhode Island's salt pond region: A special area management plan. http://www.crmc.ri.gov/regulations/SAMP_SaltPond.pdf (accessed 28 Jan. 2012).

Fonesca, M.S., W.J. Kenworthy, G.W. Thayer, NOAA Coastal Ocean Program (US) and S.F. Center. 1998. Guidelines for the conservation and restoration of seagrasses in the united states and adjacent waters. National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Southeast Fisheries Science Center, Beaufort Laboratory, [Beaufort, NC].

Koch, E.W. 2001. Beyond light: Physical, geological, and geochemical parameters as possible submersed aquatic vegetation habitat requirements. Estuaries and Coasts 24:1-17.

Lobell, D.B. and G.P. Asner. 2002. Moisture effects on soil reflectance. Soil Sci. Soc. Am. J. 66:722.

National Cooperative Soil Survey. 2012. National cooperative soil characterization database. http://ssldata.nrcs.usda.gov/ (accessed 28 Jan. 2012).

Neafsey, E.J., S.D. DeGloria, M.W. Havens, W.D. Philpot and P.J. Sullivan. 2010. Comparing Physical and Chemical Soil Properties Measurement: CNAL vs. NSSC. Soil Survey Horizons 51: 22-23.

Soil Survey Staff. 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.

U.S. Department of Agriculture, Natural Resources Conservation Service. 2012. National Soil Survey Handbook, title 430-VI. http://soils.usda.gov/technical/handbook/ (accessed 28 Jan. 2012).

USDA. 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S.D.A Agriculture Handbook #296. Natural Resources Conservation Service. United States Department of Agriculture. 669p. [http://soils.usda.gov/survey/geography/mlra, last accessed 04Jan12]

Vasques, G.M. and J.O. Sickman. 2009. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. Soil Sci. Soc. Am. J. 73:176.

Vasques, G.M., S. Grunwald and J.O. Sickman. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146:14-25.

Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics Intellig. Lab. Syst. 90:72-83.

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75.

Walvoort, D.J.J. and A.B. McBratney. 2001. Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. p. 503-507. *In* Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. ECPA, 2001.

Chapter 4: Predicting soil properties using diffuse reflectance spectroscopy of archived soil survey samples from southern New England

Abstract

Proximal diffuse reflectance spectroscopy has been proposed as a field tool to improve the speed, ability, accuracy and of soil characterization. Savings may be realized by developing soil property prediction models using archived samples to determine whether additional field sampling investments are justified. Specific study objectives are to: (1) predict soil properties using diffuse reflectance spectroscopy of archived soil survey samples from southern New England, and (2) make recommendations as to proximal diffuse reflectance spectroscopy's use in the development of soil characterization prediction models using existing soil samples. PLS1 regression models computed using reflectance spectra had RMSEP values of 9909.56 for Fe, 17.57 for % sand, 13.50 for % silt, 6.66 for % clay, 0.70 for pH, 9.34 for CEC, 1.50 for activity and 7.67 for % C; those using first-derivative spectra had RMSEP values of 9418.16 for Fe, 15.61 for % sand, 13.33 for % silt, 5.81 for % clay, 0.77 for pH, 9.46 for CEC, 1.49 for activity and 7.92 for % C. Proximal diffuse reflectance spectroscopy is recommended for use in developing soil characterization prediction models using archived soil samples.

Introduction

Proximal diffuse reflectance spectroscopy has been proposed as a field tool to improve the speed, effectiveness and accuracy of soil characterization. To realize this, a substantial commitment to develop instrument calibration standards will have to be made at the international and/or national levels. Dedicated soil survey campaigns, including the collection of soil spectra and soil sampling for laboratory

soil characterization will be required. However, most soil survey organizations, including the Natural Resources Conservation Service (NRCS), maintain collections of archived soil samples. Many of these samples have associated laboratory soil characterization data available. Given the relative economy of spectral data collection, a substantial savings can be realized by developing soil property prediction models using archived soil samples to determine whether additional field sampling investments are justified (Giasson et al., 2000). However, if these soil samples have degraded during storage, this benefit may be lost and fresh samples may need to be collected. Samples may become degraded due to the destruction of soil structure and a lack of strict climate control in storage rooms, which can potentially lead to oxidation of materials.

The goal of this study is to test the technology's previously demonstrated efficacy in estimating soil properties for archived soil samples in southern New England. These soil properties include Fe, texture (% sand, % silt and % clay), pH, cation exchange capacity (CEC), activity and % C. Previous literature suggests that proximal diffuse reflectance spectroscopy is effective in the estimation of many soil characterization properties. Cozzolino and Moron (2006) report an R^2 value of 0.90 for Fe (Cozzolino and Moron, 2003). Viscarra Rossel et al. (2006) report R^2 values of 0.59 for % sand, 0.41 for % silt, 0.60 for % clay, 0.57 for pH and 0.13 for CEC. However, Walvoort and McBratney (2001) reported a low root mean square error (RMSE) of 0.06 for % C. Estimates of soil activity have yet to be recorded in the literature, but its proportionality to clay content implies similar values. Specific objectives of this study are to: (1) predict soil properties using diffuse reflectance spectroscopy of archived soil survey samples from southern New England and (2) make recommendations as to proximal diffuse reflectance spectroscopy's use in the development of soil characterization prediction models using existing soil samples.

Materials and methods

Study area

The project study area is located within the Northeastern Forage and Forest Land Resource Region. Topography is characterized by plateaus, plains and mountains. The climate is cool and humid, experiencing an average precipitation of 865 – 1575 mm per year. Average annual temperature is 4 – 9 °C. There are 130 – 200 freeze days, with 110 in the highest elevations and 240 in some coastal areas. Total withdrawals of freshwater from the region are, on average, 72 ML/d, 93 % of which is from surface water and 7 % from groundwater. Timber and manufacturing interests consume approximately 78 % of this total, with the rest for public supply.

Soils in this region are primarily Entisols or Spodosols, commonly with a fragipan. In areas with limey geology, Alfisols occur with a fragipan. Higher elevations are characterized by Ochrept and Orthods, while lower regions feature Aqualfs, Aquepts and Histosols. Fluvents in floodplains are limited in extent but are locally important. The temperature regimes are frigid or mesic, udic in terms of soil moisture and exhibit a mixed mineralogy.

The primary land cover is forest (especially in areas of rugged relief), with 98 % of the land in private ownership. Logging, for timber and pulpwood, is a commercially important activity, as is the production of Christmas trees and maple syrup. The region grows forage and grains to feed local dairy stock, however local conditions can allow for the production of food crops for human consumption. Wildlife habitat and recreation are major land uses, especially where stoniness and steep slope has precluded other uses (USDA, 2006).

DRS measurements

Archived soil samples are stored, labeled by pedon, in cardboard boxes. Each pedon's horizons are stored within the cardboard box in a smaller cardboard tray. Each tray is labeled by horizon name and

depth interval. To provide context for the soil samples, some of the archived samples have examples of unconsolidated organic material (to illustrate organisms) and parent material (to illustrate regolith) placed before the top horizon and below the bottom horizon, respectively. Some of the archived samples have included a printout of the field description of each sample. This was used to confirm each sample's official pedon ID and horizon names. Spectra were measured with an ASD FieldSpec Pro FR spectrometer (Analytical Spectral Devices, Inc., Boulder, CO[9], using a contact probe directly against the soil profile. The spectrometer collects data at a spectral resolution of 1 nm. It is also a three-spectrometer system, with detectors for the VNIR (350 – 1000 nm), SWIR1 (1001 – 1780 nm) and SWIR2 (1781 – 2500 nm) spectral regions. Discontinuities can be observed (Figure 4.1) in the spectra in transitioning between these regions. Before and after spectral data collection for the archived soils, a dark current and a white reference (to establish zero and 100 % reflectance, respectively) (five associated spectra were stored from each) were collected to calibrate the instrument. A set of five spectra (each an instantaneous average of ten spectra) was collected from each horizon in the same manner as in the field, using a contact probe directly against the sample in the tray. A medium paintbrush was used to keep the contact probe clean between samples. Where soil samples still retained structure or were otherwise rough, effort was made to collect spectra for the smoothest available surface. It should be noted that many of the archived samples contained organic horizons (Oa, Oe and Oi), many of which were still largely unconsolidated. As a result, the collection surface necessarily has a substantial amount of relief, which may have an effect on spectral quality. Spectral samples were collected at NRCS offices in Tolland and Windsor, CT.

[9] <http://www.asdi.com>

Sampling for laboratory analysis

After profile description and spectral acquisition were complete, the Natural Resources Conservation Service (NRCS) soil survey team collected soil samples for the pedon for submission to the National Soil Survey Laboratory (NSSL) at the National Soil Survey Center (NSSC) in Lincoln, NE. All soil samples were analyzed by the NSSL to determine their physical, chemical, and mineralogical properties according to NRCS soil survey guidelines (U.S. Department of Agriculture, Natural Resources Conservation Service, 2012; National Cooperative Soil Survey, 2012).

Data analysis

To facilitate data analysis, the instrument native files (remote sensing reflectance calculated by the ASD software) for each horizon were imported into The Unscrambler software program,[10] a multivariate statistics package which specializes in spectral processing and analysis. Reflectance spectra can also be imported to universally-readable ASCII files by ASD's ViewSpecPro, but direct importation provides a more efficient workflow and ensures the preservation of variable names.

The resulting files for each horizon were averaged by a factor of five (across the five spectra collected for each horizon), yielding one spectrum for each horizon. Averaging accomplishes instrument noise reduction and accounts for variance in the samples.

First derivatives were calculated using the difference between the reflectance values associated with adjacent wavelengths at a resolution of 1 nm to preserve full spectral information. As soil moisture can reduce the amplitude of soil spectra, derivative analysis is used to difference to remove trend, highlighting the constituents of interest (ASD Inc., 2012).

[10] <http://www.camo.com/rt/Products/Unscrambler/unscrambler.html>

PLS1 regression modeling

Partial least squares regression was used to evaluate the statistical contributions of many X-variables (predictors) to one Y-variable (responses). In this study, the reflectance values for the waveband 350 – 2500 nm comprised the set of X-variables. Excluded due to falloff in detector sensitivity were the wavebands 350 – 450 nm and 2400 – 2500 nm (marked in gray in Figure 4.1). Excluded also were water absorption wavebands 1350 – 1500 nm and 1850 – 2000 nm (marked in blue in Figure 4.1). Selected physical and chemical soils properties were the Y-variables. A common statistical tool in soil chemometrics, PLS1 regression allows the user to determine which spectral wavelengths are correlated with properties of interest (Vasques et al., 2008; Vasques and Sickman, 2009; Viscarra Rossel, 2008).

Results and discussion

Sample descriptions

Full data descriptions of archived soil samples are located online at the National Cooperative Soil Survey's Soil Characterization Data website[11] and summarized here for selected soil properties. According to the median (Table 4.1), the soils in this study are characterized by moderate to high Fe content, as well as high sand content, with relatively low silt and clay contents. They are also very strongly acid, have a CEC outside of general agricultural limits and are slightly active. Carbon content is also relatively low.

They are relatively high-reflectance soils (Figure 4.1), represented here by a single pedon of a Hollis series soil, having a maximum reflectance of approximately 0.45. Particularly noticeable are the low reflectances of the organic O-horizons and the high reflectances of the mineral B-horizons in the visible wavebands (400 – 700 nm). Water absorption bands (marked in blue; 1350 – 1500 nm and 1850 – 2000 nm) are relatively shallow, but present, indicating the samples are retaining ambient moisture in office

[11] <http://ssldata.nrcs.usda.gov/>

storage conditions. In the first-derivative plot, Bw1 and Bw2 look very similar throughout the spectrum. Horizon A tracks them well except in the blue and green. Horizons Oe and Oi also look similar to one another throughout the spectrum.

	Lab - Fe	Lab - % Sand	Lab - % Silt	Lab - % Clay	Lab - pH	Lab – CEC	Lab - Activity	Lab - % C
Units	mg/kg	%	%	%	pH	meq+/100g	scalar	%
#	96	253	253	247	289	303	242	303
Mean	31308.40	55.56	36.33	8.27	5.11	8.04	1.69	9.36
Max	67265.00	99.30	84.20	56.70	8.40	77.81	12.39	53.86
Min	4609.00	1.30	0.50	0.10	3.20	0.00	0.24	0.01
Range	62656.00	98.00	83.70	56.60	5.20	77.81	12.15	53.85
Std Deviation	14618.21	20.95	17.22	7.25	0.99	10.76	1.67	15.37
Variance	213692100.00	438.82	296.50	52.60	0.99	115.85	2.80	236.10
RMS	34520.75	59.36	40.19	10.99	5.21	13.42	2.38	17.97
Skewness	0.35	-0.27	0.39	3.33	1.19	2.69	2.82	1.76
Kurtosis	-0.53	0.01	0.21	16.54	1.35	10.01	11.15	1.66
Median	28590.50	56.20	35.50	7.50	4.90	4.70	1.09	1.52
Q1	21153.00	44.80	26.30	4.10	4.50	0.78	0.65	0.40
Q3	43355.75	69.00	45.20	10.60	5.60	10.16	2.00	7.46

Table 4.1 – Soil characterization properties summary statistics

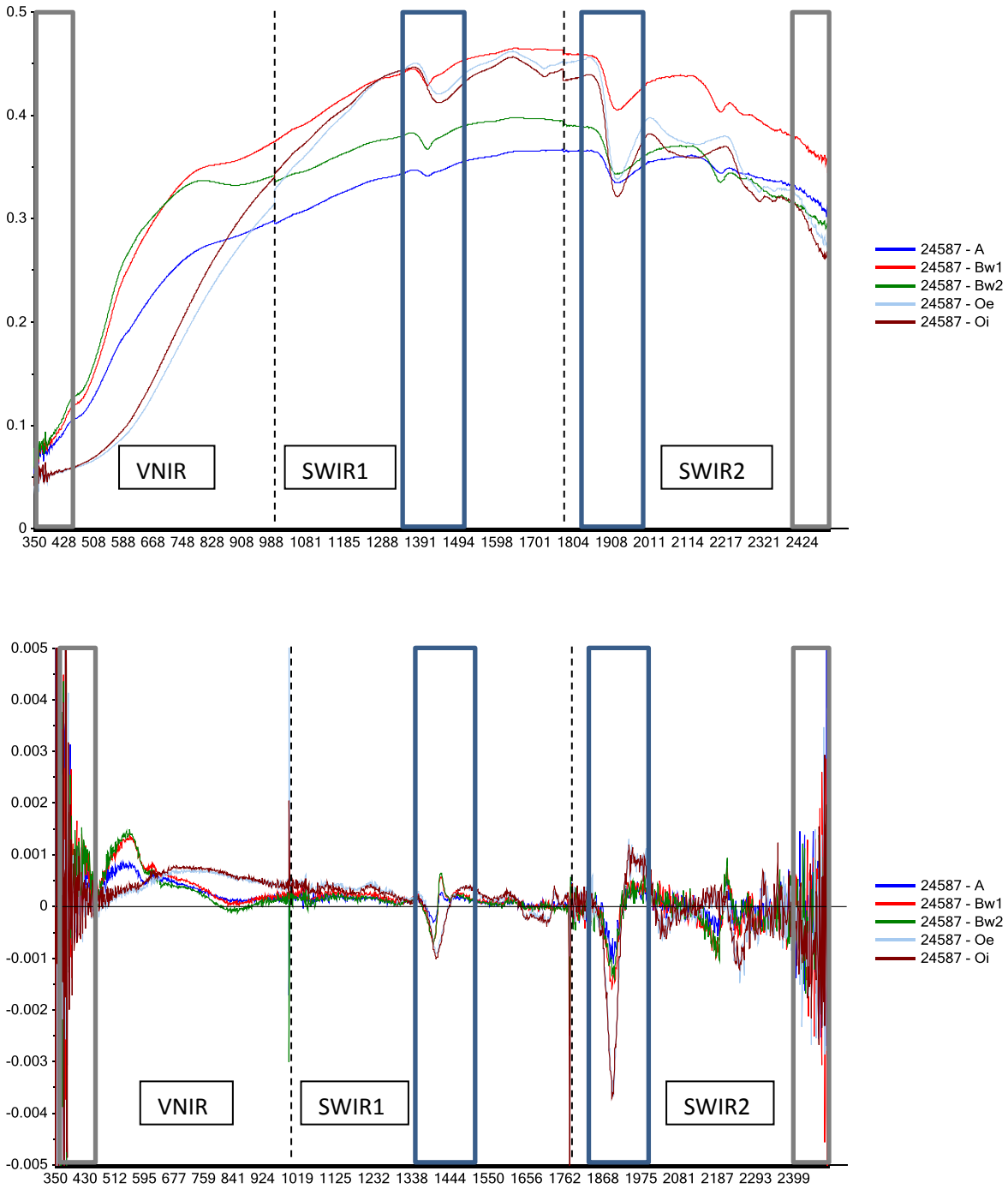


Figure 4.1 – Sample spectra reflectance and first-derivative (Hollis soil)

Soil characterization PLS1 models

Statistical analyses conducted for selected soil physical and chemical properties in the following tables

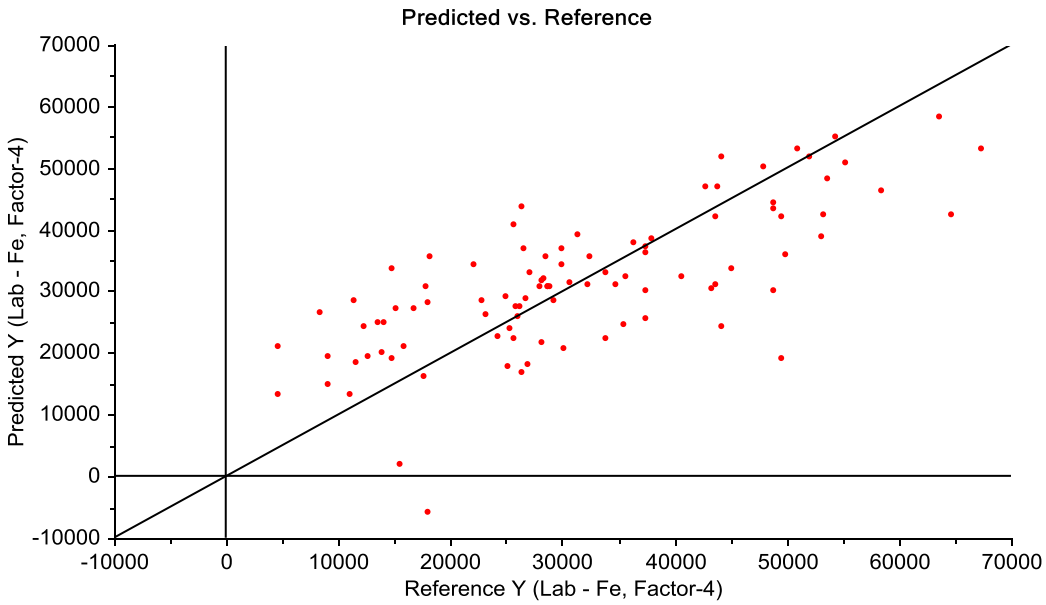
and plots. Validation PLS1 regression values are presented by spectral pre-treatment (reflectance and

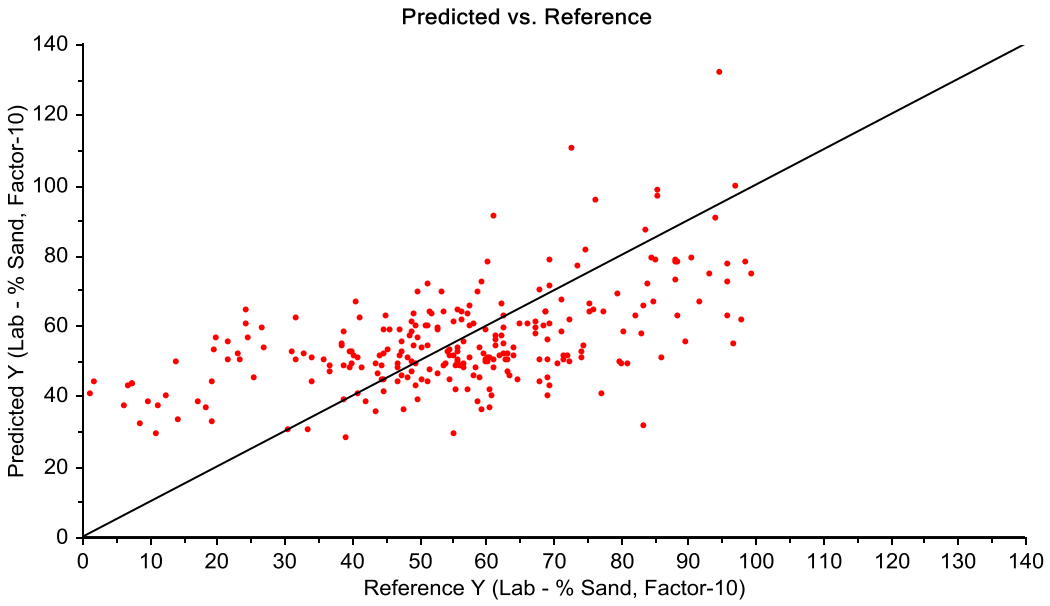
first-derivative) (Tables 4.2 – 4.3). Included are the root-mean-squared errors of prediction (RMSEP), the trendline coefficients in the form $y = Slope \times x + Offset$ and correlation values. In addition, plots showing the measured versus the predicted soil properties values are included for the reflectance and first-derivative data (Figures 4.2 and 4.3).

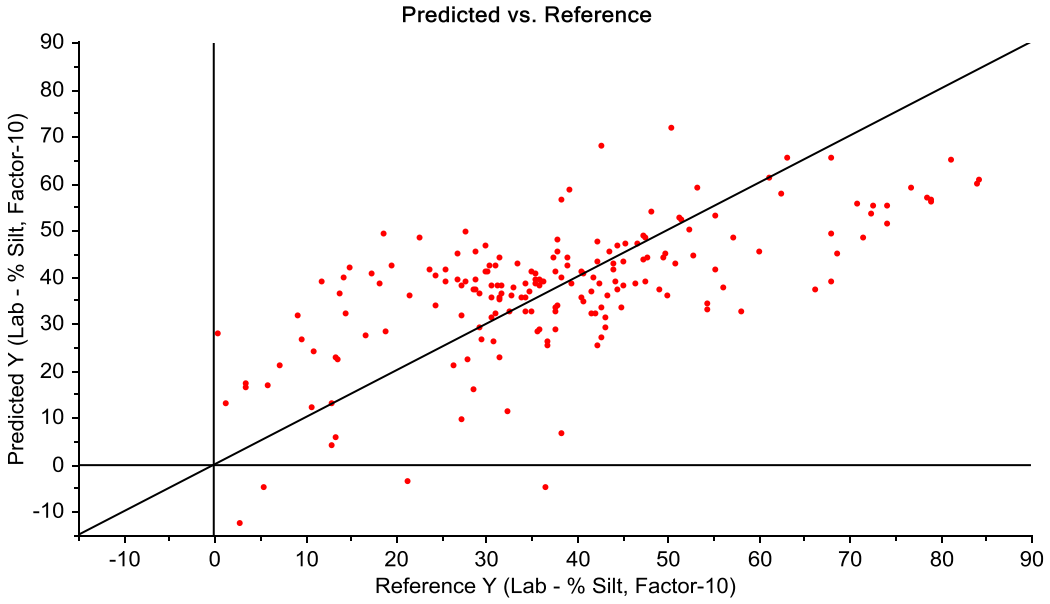
Reflectance

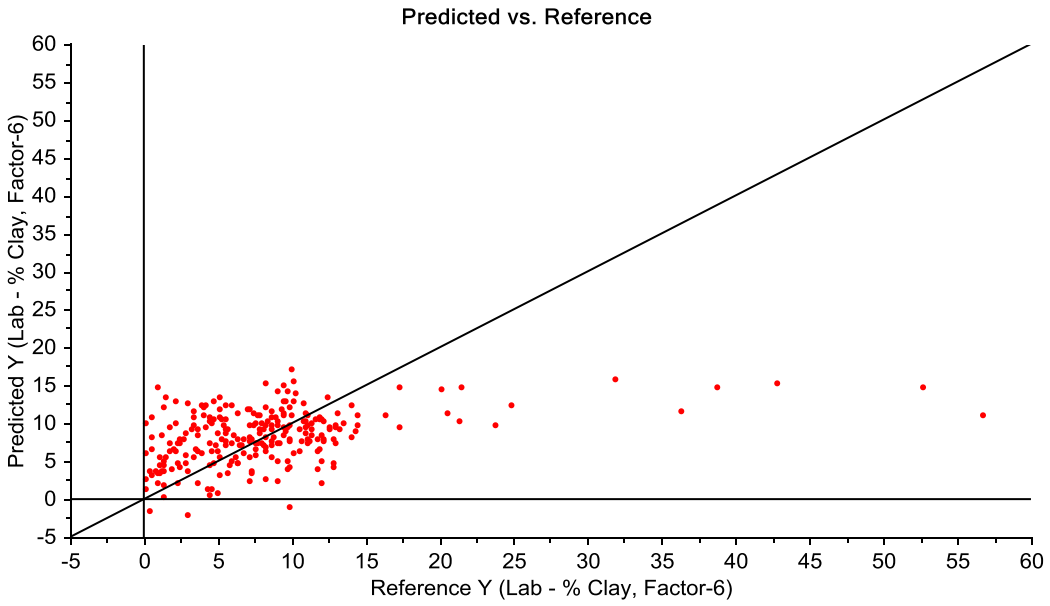
Property	RMSEP	Slope	Offset	Correlation
Fe	9909.56	0.57	13497.72	0.73
% Sand	17.57	0.37	35.03	0.55
% Silt	13.50	0.50	18.75	0.65
% Clay	6.66	0.20	6.67	0.40
pH	0.70	0.64	1.83	0.73
CEC	9.34	0.27	5.87	0.50
Activity	1.50	0.21	1.33	0.44
% C	7.67	0.75	2.29	0.87

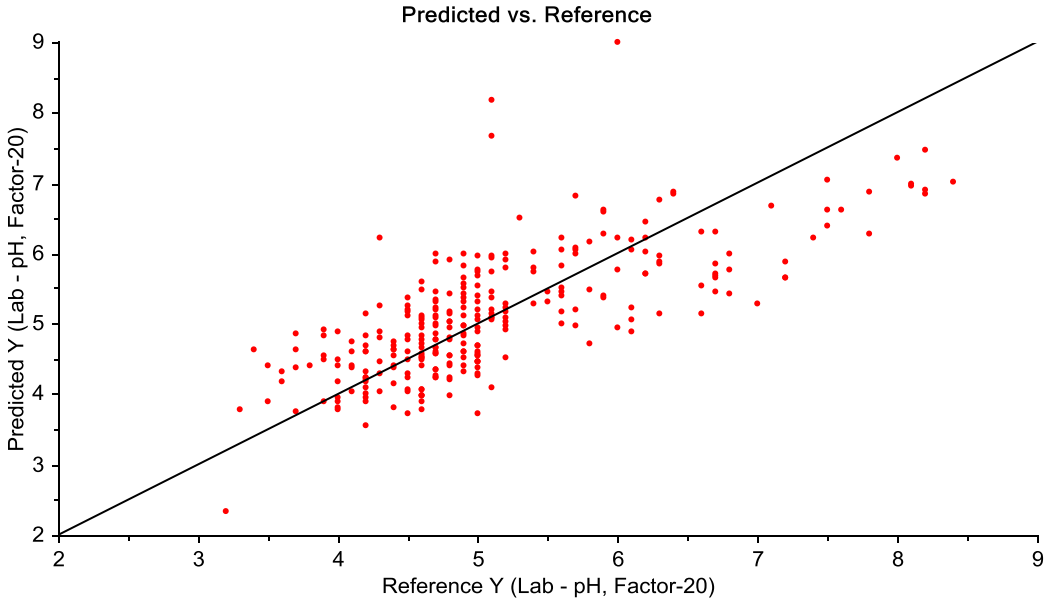
Table 4.2 – Prediction diagnostics for soil characterization PLS1 models (reflectance)

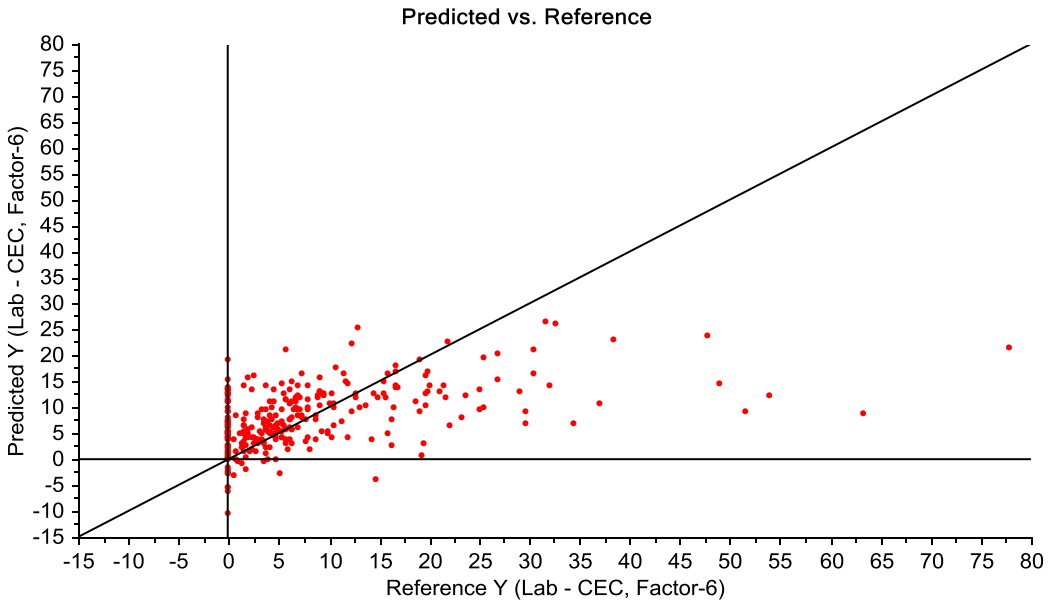


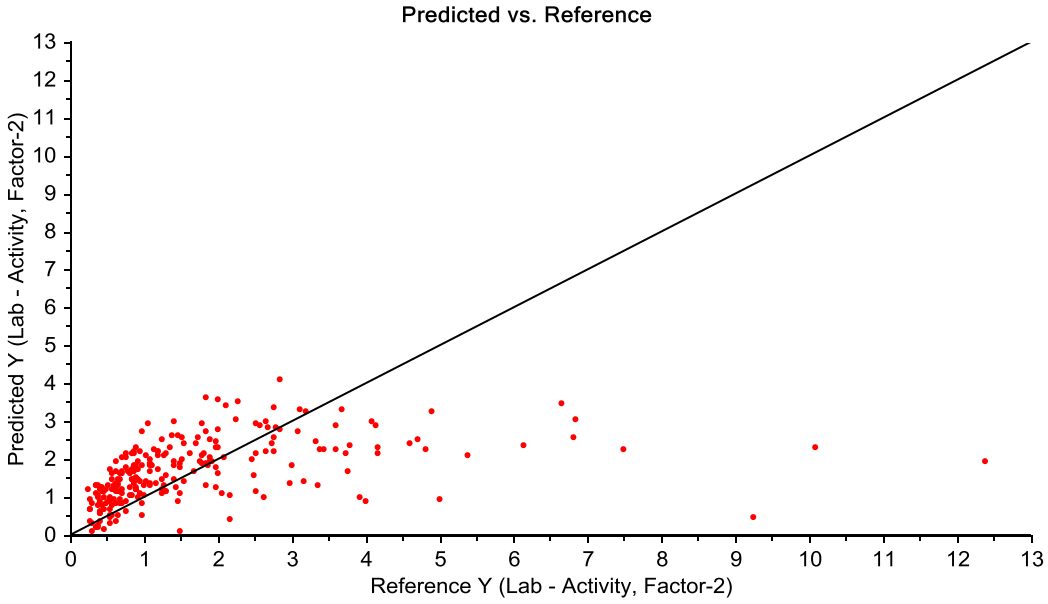

Predicted vs. Reference
Predicted Y (Lab - % Sand, Factor-10)
Reference Y (Lab - % Sand, Factor-10)
0
20
40
60
80
100
120
140
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140


Predicted vs. Reference
Predicted Y (Lab - % Silt, Factor-10)
Reference Y (Lab - % Silt, Factor-10)
-10
0
10
20
30
40
50
60
70
80
90


Predicted vs. Reference
Predicted Y (Lab - % Clay, Factor-6)
Reference Y (Lab - % Clay, Factor-6)
60
55
50
45
40
35
30
25
20
15
10
5
0
-5
-5
0
5
10
15
20
25
30
35
40
45
50
55
60


Predicted vs. Reference
Predicted Y (Lab - pH, Factor-20)
Reference Y (Lab - pH, Factor-20)
9
8
7
6
5
4
3
2
2
3
4
5
6
7
8
9


Predicted vs. Reference
Predicted Y (Lab - CEC, Factor-6)
Reference Y (Lab - CEC, Factor-6)


Predicted vs. Reference
Predicted Y (Lab - Activity, Factor-2)
Reference Y (Lab - Activity, Factor-2)

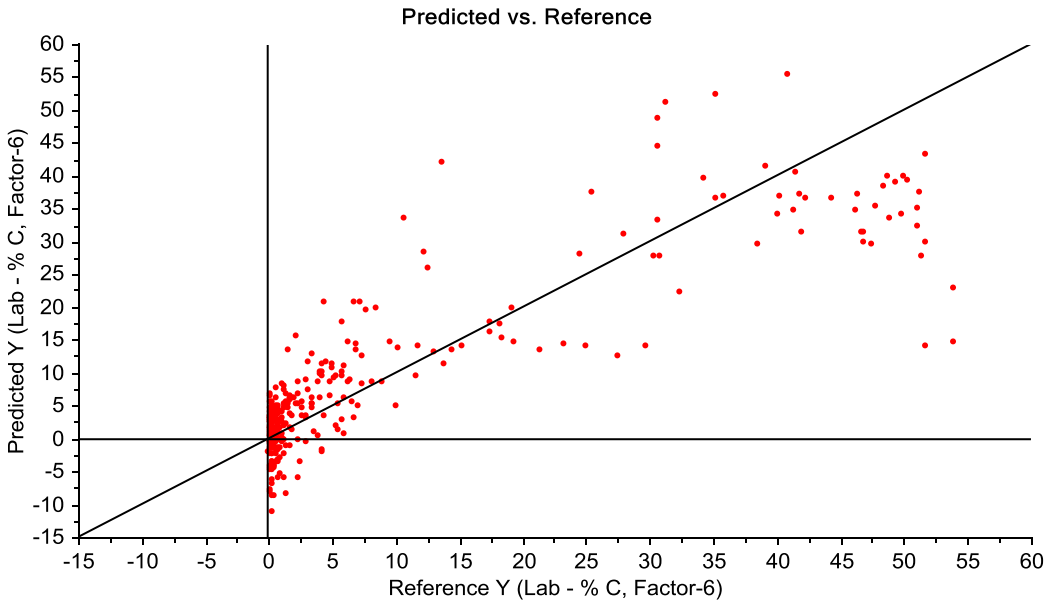
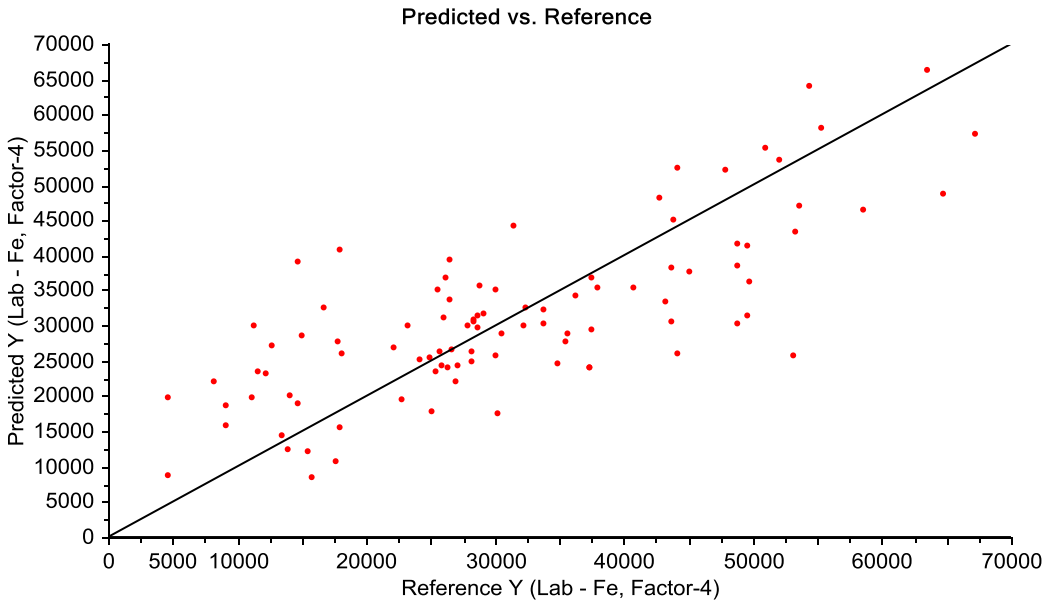


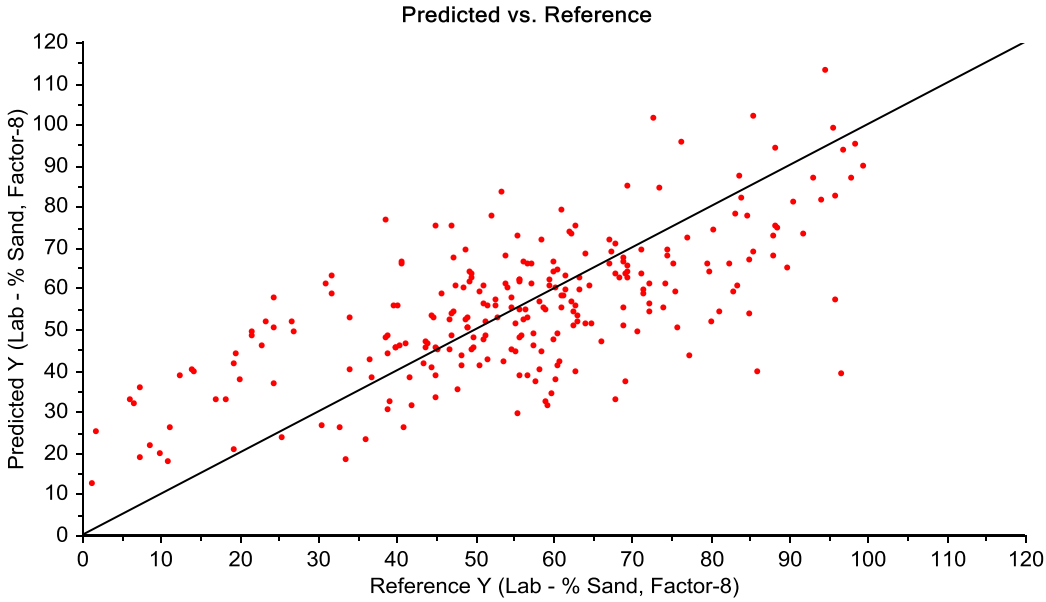
Figure 4.2 – Predicted vs. measured for soil characterization properties (reflectance)

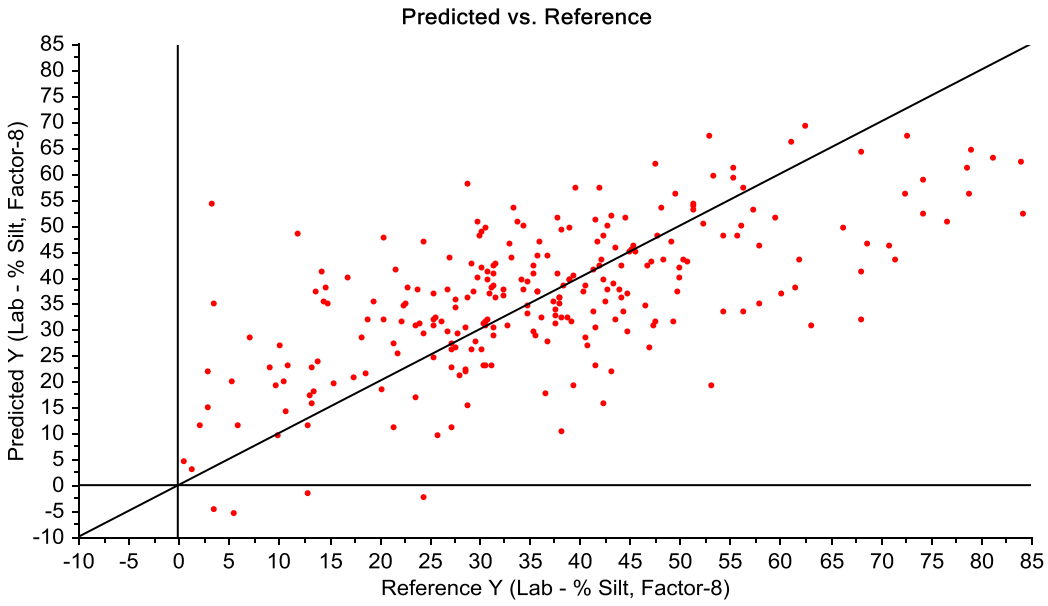
First-derivative

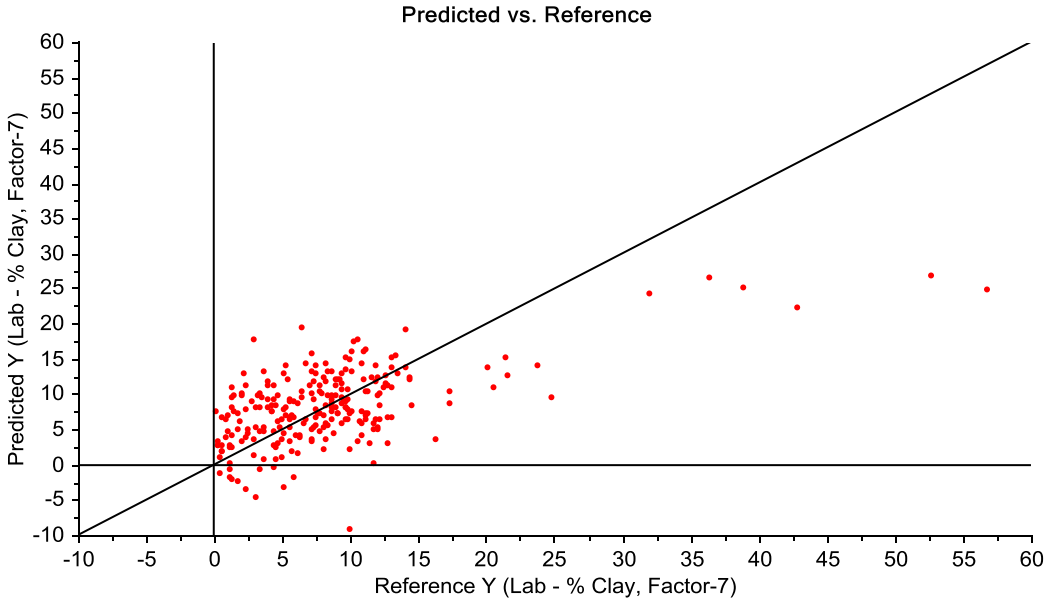
Property	RMSEP	Slope	Offset	Correlation
Fe	9418.16	0.61	12095.66	0.76
% Sand	15.61	0.54	25.08	0.68
% Silt	13.33	0.52	17.64	0.65
% Clay	5.81	0.44	4.55	0.61
pH	0.77	0.51	2.54	0.65
CEC	9.46	0.28	5.70	0.48
Activity	1.49	0.24	1.28	0.46
% C	7.92	0.75	2.34	0.86

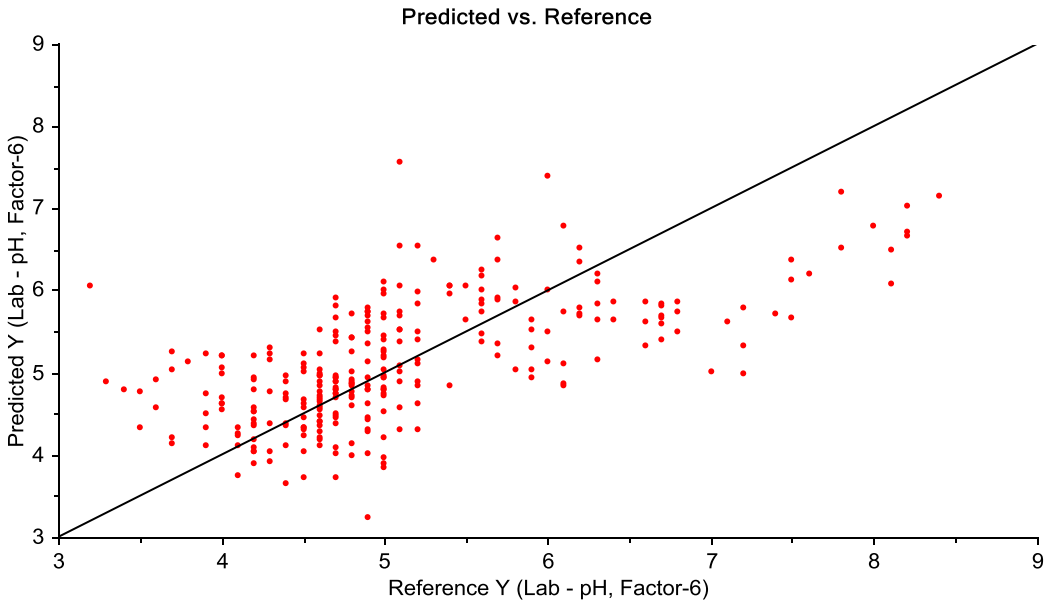
Table 4.3 – Prediction diagnostics for soil characterization PLS1 models (first-derivative)

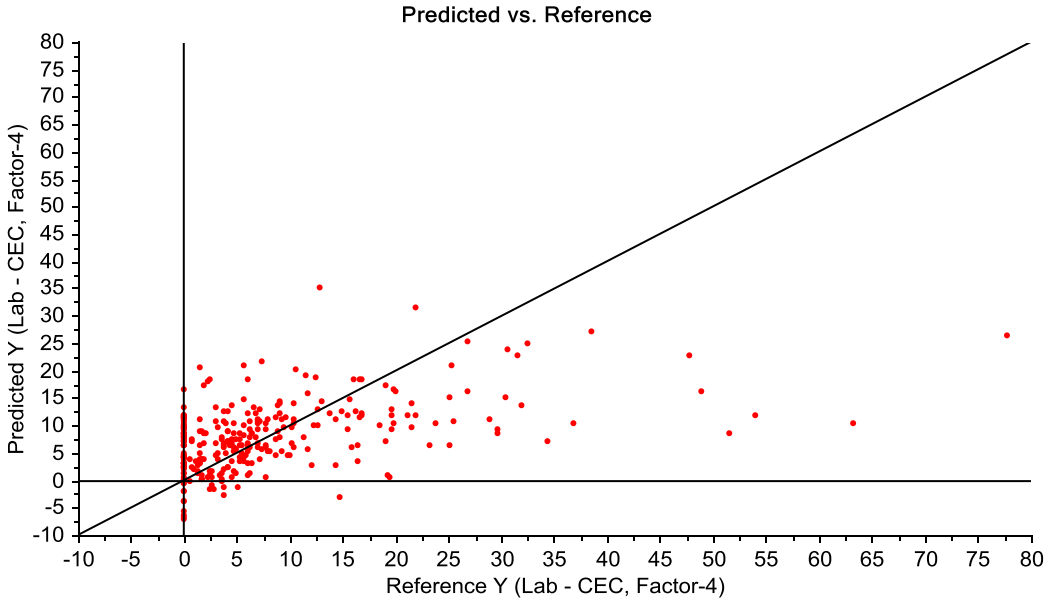

Predicted vs. Reference
Predicted Y (Lab - Fe, Factor-4)
Reference Y (Lab - Fe, Factor-4)
0
5000
10000
15000
20000
25000
30000
35000
40000
45000
50000
55000
60000
65000
70000


Predicted vs. Reference
Predicted Y (Lab - % Sand, Factor-8)
Reference Y (Lab - % Sand, Factor-8)
0
10
20
30
40
50
60
70
80
90
100
110
120


Predicted vs. Reference
Predicted Y (Lab - % Silt, Factor-8)
Reference Y (Lab - % Silt, Factor-8)


Predicted vs. Reference
Predicted Y (Lab - % Clay, Factor-7)
Reference Y (Lab - % Clay, Factor-7)


Predicted vs. Reference
Predicted Y (Lab - pH, Factor-6)
Reference Y (Lab - pH, Factor-6)
3
4
5
6
7
8
9


Predicted vs. Reference
Predicted Y (Lab - CEC, Factor-4)
Reference Y (Lab - CEC, Factor-4)
-10
-5
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80

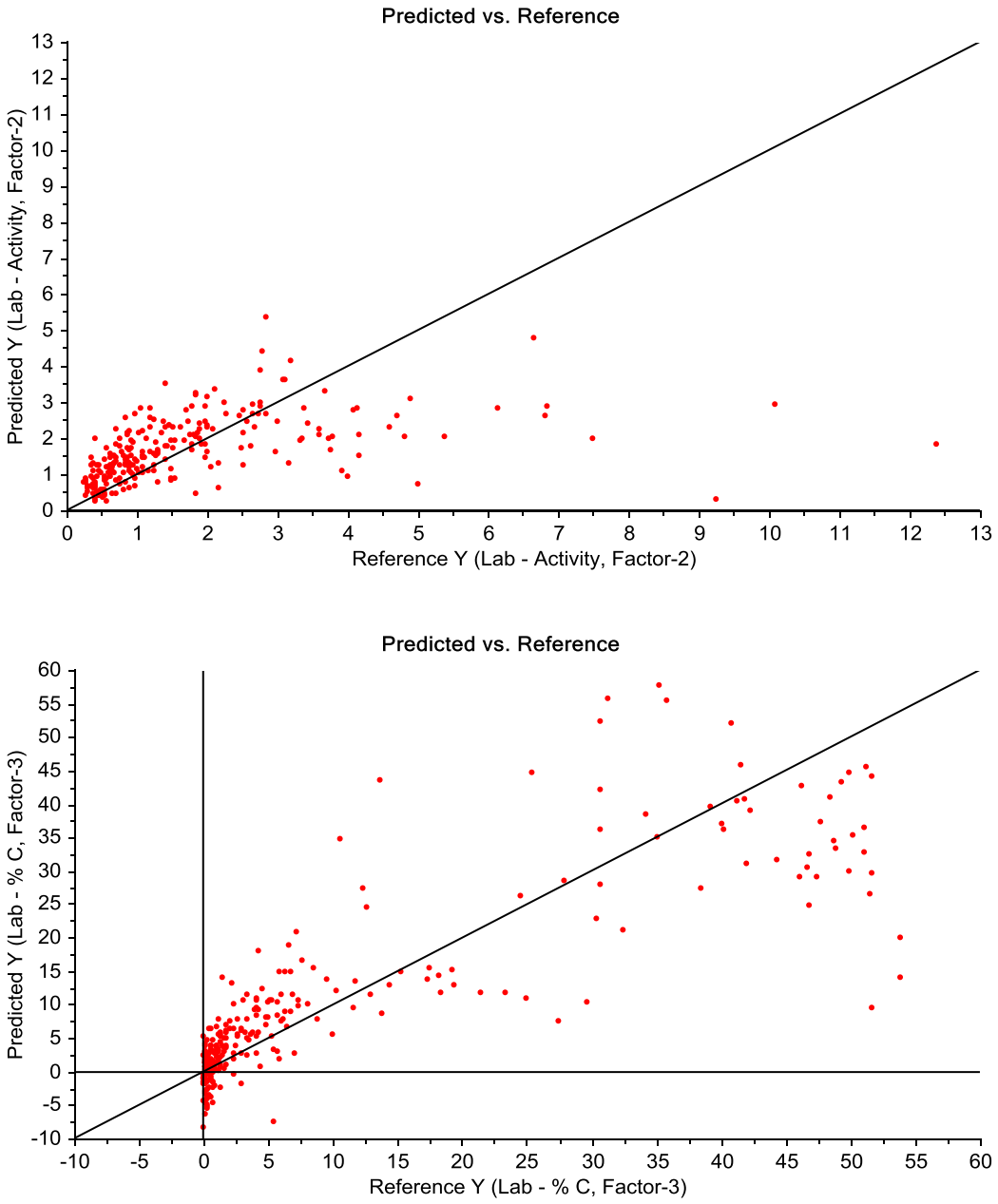


Figure 4.3 – Predicted vs. measured for soil characterization properties (first-derivative)

Interpretations

Since the RMSEP for each property (Tables 4.2 and 4.3) falls within the standard deviation of reference values for the same (Table 4.1), proximal diffuse reflectance spectroscopy can be recommended for use in characterizing archived soil samples. The standard deviation of Fe is 14 618.21, while the RMSEP for reflectance models is 9909.56 and 9418.16 for first-derivative models. The standard deviation of % sand

is 20.95, while the RMSEP for reflectance models is 17.57 and 15.61 for first-derivative models. The standard deviation of % silt is 17.22, while the RMSEP for reflectance models is 13.50 and 13.33 for first-derivative models. The standard deviation of % clay is 7.25, while the RMSEP for reflectance models is 6.66 and 5.81 for first-derivative models. The standard deviation of pH is 0.99, while the RMSEP for reflectance models is 0.70 and 0.77 for first-derivative models. The standard deviation of CEC is 10.76, while the RMSEP for reflectance models is 9.34 and 9.46 for first-derivative models. The standard deviation of activity is 1.67, while the RMSEP for reflectance models is 1.50 and 1.49 for first-derivative models. The standard deviation of % C is 15.37, while the RMSEP for reflectance models is 7.67 and 7.92 for first-derivative models. Reflectance models performed better for pH, CEC and % C. First-derivative models performed better for Fe, % sand, % silt, % clay and activity.

Percent C was well predicted, even though most of the samples had low values, but with a high range of 54. As with field benchmark samples, % sand is over predicted for low values and underpredicted for high values. Predicted % silt values generally follow the target line in Figure 4.2. The high number of low % clay values may have degraded model performance. This also holds for pH, which always has a narrow numeric range and additionally is an indirectly sensed soil property. CEC and activity are highly underpredicted. Given the possibility of sample degradation and gaps in soil characterization data, it is highly possible that proximal diffuse reflectance spectroscopy can be suitable for these properties. Additionally, the heterogeneity of the soil samples may account for deficiencies in model performance. In many of the soils collected in eastern Connecticut, pedons are comprised of highly organic O-horizons above Bw horizons with heavy mineral content, yielding a wide range in characterization values within a single profile.

Statistically, Figures 4.2 and 4.3 indicate that while a relatively low number of principal components (2 – 20) were required to optimize model performance. Additionally, models computed using first-derivative

spectra required fewer principal components (2 – 8), than those computed using reflectance spectra (2 – 20). The normalizing effect of the first-derivative may have improved the quality of the predictors, yielding an optimized model at earlier principal components. In terms of point scatter, the plots reflect the underlying skewness and kurtosis (Table 4.1) of the reference data. For example, % clay, CEC, activity and % C have high levels of skewness (3.33, 2.69, 2.82 and 1.76 respectively) and kurtosis (16.54, 10.01, 11.15, 1.66 respectively), leading to a high number of samples clustering near the prediction axis and underpredicted values for samples with reference values at the top end of their range.

Summary and conclusion

Proximal diffuse reflectance spectroscopy is recommended for use with archived soil samples, subject to the limitations described above. Practitioners will need to evaluate published RMSEP values within the context of soil taxonomic requirements (Soil Survey Staff, 2010). Since verbal taxonomic designations have a quantitative basis, errors in prediction could yield substantially different classifications. However, as reference soil characterization values for the same samples can vary by between major laboratories (Neafsey et al., 2010) errors in classification may be present using conventional techniques.

The soil samples archived at NRCS offices at the state level offer a wealth of information which can help form the foundation of soil spectral infrastructure. Given sample heterogeneity, official prediction models should be calculated within soil series. The emerging role of MLRAs in NRCS soil survey activities suggests that this may be a desirable basis for coordinating soil spectra collection and increase sample size within applicable soil classes. Additionally, it may be necessary to develop prediction models for horizon classes due to material differences within the soil profile.

Central to the development of archived soil samples for spectral data collection is the adoption of standardized procedures. The role of the NSSL in coordinating MLRA offices' execution of the Rapid

Carbon Assessment project[12] may be a good starting point in achieving this goal. The leadership of the NSSL in collecting soil spectra for their archived holdings will continue to be instrumental. The efficacy of proximal diffuse reflectance spectroscopy relies on large sample sizes for a diversity of soils and NRCS' archived samples are one of several means (including land grant and other agricultural institutions' holdings) to this end.

[12] <ftp://ftp-fc.sc.egov.usda.gov/NSSC/NCSS/Conferences/regional/2010/north_central/west2.pdf>

References

ASD Inc. 2012. What is a derivative spectrum? http://www.asdi.com/resource-center/faqs/what-is-a-derivative-spectrum (accessed 28 Jan. 2012).

Cozzolino, D. and A. Moron. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. The Journal of Agricultural Science 140:65-71.

Giasson, E., C. van Es, A. van Wambeke and R.B. Bryant. 2000. Assessing the economic value of soil information using decision analysis techniques. Soil Sci. 165:971.

National Cooperative Soil Survey. 2012. National cooperative soil characterization database. http://ssldata.nrcs.usda.gov/ (accessed 28 Jan. 2012).

Neafsey, E.J., S.D. DeGloria, M.W. Havens, W.D. Philpot and P.J. Sullivan. 2010. Comparing Physical and Chemical Soil Properties Measurement: CNAL vs. NSSC. Soil Survey Horizons 51: 22-23.

Soil Survey Staff. 2010. Keys to Soil Taxonomy, 11th ed. USDA-Natural Resources Conservation Service, Washington, DC.

U.S. Department of Agriculture, Natural Resources Conservation Service. 2012. National Soil Survey Handbook, title 430-VI. http://soils.usda.gov/technical/handbook/ (accessed 28 Jan. 2012).

USDA. 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S.D.A Agriculture Handbook #296. Natural Resources Conservation Service. United States Department of Agriculture. 669p. [http://soils.usda.gov/survey/geography/mlra, last accessed 04Jan12]

Vasques, G.M. and J.O. Sickman. 2009. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. Soil Sci. Soc. Am. J. 73:176.

Vasques, G.M., S. Grunwald and J.O. Sickman. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146:14-25.

Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. Chemometrics Intellig. Lab. Syst. 90:72-83.

Viscarra Rossel, R.A., D.J.J. Walvoort, A.B. McBratney, L.J. Janik and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131:59-75.

Viscarra Rossel, R.A., Y.S. Jeon, I.O.A. Odeh and A.B. McBratney. 2008. Using a legacy soil sample to develop a mid-IR spectral library. Aust. J. Soil Res. 46:1-16.

Walvoort, D.J.J. and A.B. McBratney. 2001. Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. p. 503-507. *In* Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. ECPA, 2001.

Chapter 5: Summary and conclusion

Significant findings

This dissertation has met its objectives to:

Develop a design for a comprehensive visible-near infrared diffuse reflectance spectroscopy (VNIR DRS) spectral library and PLS1 regression prediction models for some of southern New England's benchmark soils and archived soil survey samples

- Proposed VNIR DRS spectral library
    - Distributed database for soil chemometric and other applications
    - Geodatabase for spatial pedometrics
- PLS1 regression prediction models
    - Benchmark soils
        - Models computed using reflectance spectra had root-mean-squared errors of prediction (RMSEP) values of 18.89 for % sand, 14.83 for % silt, 5.00 for % clay, 0.75 for pH, 4.95 for cation exchange capacity (CEC), 1.84 for activity and 11.84 for % C; those using first-derivative spectra had RMSEP values of 21.62 for % sand, 20.21 for % silt, 4.80 for % clay, 0.77 for pH, 4.74 for CEC, 1.87 for activity and 10.52 for % C
        - VNIR DRS is recommended for use with benchmark soils
    - Archived soil survey samples
        - For archived soils, models computed using reflectance spectra had RMSEP values of 9909.56 for Fe, 17.57 for % sand, 13.50 for % silt, 6.66 for % clay, 0.70 for pH, 9.34 for CEC, 1.50 for activity and 7.67 for % C; those using first-derivative spectra had RMSEP values of 9418.16 for Fe, 15.61 for % sand, 13.33

for % silt, 5.81 for % clay, 0.77 for pH, 9.46 for CEC, 1.49 for activity and 7.92 for % C

- VNIR DRS is recommended for use with archived soils

Support subaqueous soil mapping efforts in southern New England estuaries and selected salt ponds in Rhode Island by calibrating VNIR DRS as a covariate

- For subaqueous soils, models computed using reflectance spectra had RMSEP values of 25.81 for % sand, 22.94 for % SOC, 15.06 for salinity and 0.76 for pH; those using first-derivative spectra had RMSEP values of 26.01 for % sand, 20.63 for % SOC, 12.17 for salinity and 0.72 for pH
- VNIR DRS is recommended for use with subaqueous soils

Determine the utility of using VNIR DRS for measuring sulfidic materials in subaqueous soils

- For subaqueous soils, the total sulfide model computed using reflectance spectra had a RMSEP value of 1355.72 for total sulfide; using first-derivative spectra it had a RMSEP value of 1317.72 for total sulfide
- Combined with good results for pH, it is likely that VNIR DRS will be a useful management tool in dealing with sulfidic materials in the built environment

Provide technical assistance to National Cooperative Soil Survey (NCSS) soil survey programs in southern New England

- Participated in benchmark soil survey operations during the 2008 field season
- Demonstrated VNIR DRS capability with mesic soil types
- Collected spectra for described subaqueous soils
- Collected spectra for NRCS-CT's archived soil samples
- Worked with the NCSS to ensure the compatibility of collected spectra with those being collected at the national level

Recommendations for future research

Spectral library development

1. Develop strong international standards for acquiring soil spectra
2. Collect spectra for all available samples present in the National Cooperative Soil Characterization Database (NCSCD)
3. Support multiple spectral prediction model types
4. Expand proximal sensing capabilities
5. Include proximal soil sensing in soil science curricula
6. Leverage existing soil survey partnerships

VNIR DRS as a pedometric platform

While inquiry must continue to enhance interpretation of VNIR DRS-acquired spectra with respect to soil chemistry and physics, it is time that VNIR DRS was developed for field and laboratory pedometric applications. There is a high degree of interest in these uses within the soil science community and associated stakeholder groups in seeing this come to fruition. Ongoing work on imaging pedometric spectrometers continue to be encouraging and hold promise both as a soil science and mapping tool. Basic matter-energy research continues with respect to our remote sensing platforms, while also supporting users in answering their science questions. This model can be applied to pedometric VNIR DRS as well, concurrent with the development of spectral libraries. As in remote sensing, users of VNIR DRS will expand the quantity of data available to answer key questions about how VNIR DRS works. To accomplish these ends, the following steps are recommended:

1. Establish soil characterization prediction data quality standards
2. Create a validated spectral library
3. Develop and archive prototype predictive models for key soil properties

4. Program a software application which can calculate, store and display predicted soil characterization data using spectra from soils samples of unknown content
5. Convene VNIR DRS pedometric user groups regularly to ensure quality of product

Appendix

Lab characterization data

Full lab characterization data can be obtained at http://ssldata.nrcs.usda.gov/advquery.asp, using the multiple list option (Figure A.1) with the Pedons(Lab#) as the set criteria as shown below:

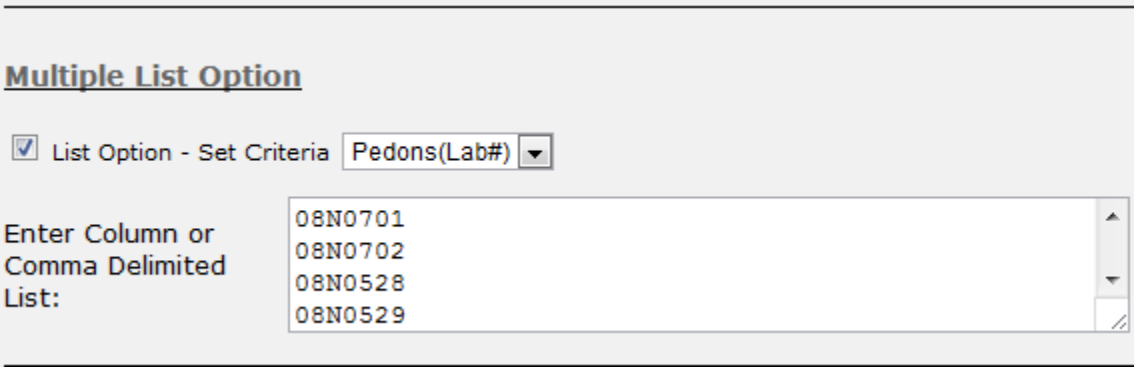


Figure A.1 - NCSS Soil Characterization Advanced Query Interface (Multiple List Option)

Field	Subaqueous	Subaqueous*	Lab
08N0697	06N0547	04RI009003	99P0553
08N0698	06N0549	05RI009006 [A-NP]	99P0554
08N0699	06N0550	05RI009011 [A-NP]	99P0556
08N0700	06N0551	05RI009012 [A-NP]	99P0557
08N0701	07N0037	06RI003003 [GB]	99P0558
08N0702	07N0038	06RI003004 [GB]	99P0559
08N0528	07N0040	06RI003007 [GB]	99P0560
08N0529	07N0041	06RI003008 [GB]	00P0198
	07N0044	06RI009003 [WH]	00P0200
		06RI009004 [WH]	00P0202
		06RI009005 [WH]	00P0203
		06RI009007 [WH]	00P0204
		06RI009013 [QP]	01N0944
		06RI009014 [QP]	01N0945
		06RI009015 [QP]	01N0946
		06RI009016 [QP]	01N0947
		06RI009017 [QP]	01N0948
		06RI009018 [QP]	01N0949
		07RI009001 [QP]	01N0950
		07RI009003 [QP]	01N0951
		07RI009006 [PJ]	01N0952

Field	Subaqueous	Subaqueous*	Lab
		07RI009007 [PJ]	01N0953
		07RI009009 [PJ]	02N1096
		07RI009010 [PJ]	02N1098
		07RI009011 [PJ]	02N1099
		07RI009012 [PJ]	02N1100
		07RI009013 [PJ]	02N1101
			02N1102
			02N1103
			02N1104
			02N1105
			05N0109
			05N0110
			05N0111
			04N1029
			04N1030
			04N1031
			04N1032
			04N1033
			04N1034
			04N1035
			05N0877
			05N0878
			05N0879
			05N0882
			05N0883
			05N0884
			05N0885
			07N0026
			07N0027
			07N0028

* Characterization data for these pedons can be found at http://nesoil.com/sasdata/

Table A.1 – Lab characterization data pedon IDs

Spectral data

Spectral data used in this dissertation will be made available on the NRCS website and searchable using the pedon IDs in the previous section. Instructions on data access will be published on https://sites.google.com/site/ejneafsey/weeklyprogressreports when these data are made public.

Sampling locations

Sample locations used in this dissertation will be published in KML format. Instructions on data access will be published on https://sites.google.com/site/ejneafsey/weeklyprogressreports when these data are made public.

References

National Cooperative Soil Survey. 2012. National cooperative soil characterization database.

http://ssldata.nrcs.usda.gov/ (accessed 28 Jan. 2012).